Multiply Robust Causal Mediation Analysis with Continuous Treatments

AmirEmad Ghassami*, Numair Sani*, Yizhen Xu*, Ilya Shpitser

Abstract

In many applications, researchers are interested in the direct and indirect causal effects of an intervention on an outcome of interest. Mediation analysis offers a rigorous framework for the identification and estimation of such causal quantities. In the case of binary treatment, efficient estimators for the direct and indirect effects are derived by Tchetgen Tchetgen and Shpitser (2012). These estimators are based on influence functions and possess desirable multiple robustness properties. However, they are not readily applicable when treatments are continuous, which is the case in several settings, such as drug dosage in medical applications. In this work, we extend the influence function-based estimator of Tchetgen Tchetgen and Shpitser (2012) to deal with continuous treatments by utilizing a kernel smoothing approach. We first demonstrate that our proposed estimator preserves the multiple robustness property of the estimator in Tchetgen Tchetgen and Shpitser (2012). Then we show that under certain mild regularity conditions, our estimator is asymptotically normal. Our estimation scheme allows for high-dimensional nuisance parameters that can be estimated at slower rates than the target parameter. Additionally, we utilize cross-fitting, which allows for weaker smoothness requirements for the nuisance functions.

1 Introduction

Estimating the effect of a treatment, policy or intervention is of interest in various fields such as epidemiology, economics, medicine and sociology. A common estimand is the Average Causal Effect (ACE), which has been extensively studied in the literature (Hernán and Robins, 2020). However, in addition to estimating the treatment effect, one can also be interested in the pathways and mechanisms through which the treatment affects the outcome of interest. Causal mediation analysis offers a precise and rigorous mathematical framework to answer such questions. Causal mediation analysis has been explored in depth in the literature, see (Robins and Greenland, 1992; Tchetgen Tchetgen and Shpitser, 2012; Pearl,

$*$ denotes equal contribution.

2001; VanderWeele, 2009; Goetgeluk et al., 2008; Imai et al., 2010; van der Laan and Petersen, 2008; Lange and Hansen, 2011; Lange et al., 2012).

Much of the literature on mediation analysis assumes binary treatments. However, interventions involving the dosage of a drug, the duration of an activity, or the frequency of activity are better described with continuous treatments. In such cases, mediation effects are naturally described by a multi-dimensional surface rather than a scalar parameter. This scenario is challenging because it involves learning a multi-dimensional surface without imposing a priori shape constraints. Additionally, in the presence of high dimensional nuisance parameters, the estimator may inherit the slow rates of our nuisance estimators, adversely affecting inference for the target parameter.

The challenges related to estimating ACE in the continuous treatment setting have been addressed by (Kennedy et al., 2017; Ai et al., 2021; Hirano and Imbens, 2004; Kreif et al., 2015; Imbens, 2000; Su et al., 2019; Kallus and Zhou, 2018; Colangelo and Lee, 2020; Hill, 2011). A common method for dealing with continuous treatments involves using Bayesian Additive Regression Trees (BART) as used by Hill (2011). However, this requires correct specification of the relevant models, and inherits the rate of the outcome regression estimation. An alternative approach that leverages semiparametric theory involves specifying a parametric form for the dose response curve, or projecting the true curve onto a parametric model as presented by (Robins, 2000; Van Der Laan and Robins, 1998; Neugebauer and van der Laan, 2007). However, these methods may suffer from bias under misspecification of the dose response curve.

In contrast to approaches involving parametric assumptions on the dose response curve, flexible approaches to modeling the dose response curve have also been proposed. For example, Kennedy et al. (2017) utilize a two-stage estimator that first constructs a doubly robust pseudo-outcome in the first stage and then regresses the pseudo-outcome on the treatment in the second-stage using non-parametric regression methods. Colangelo and Lee (2020) utilize double machine learning along with applying kernel smoothing to the Augmented Inverse Propensity Weighted (AIPW) score (Robins and Rotnitzky, 1995). This allows for slower estimation of nuisance parameters while still obtaining fast rates for the target parameter. However, dealing with continuous treatments in mediation analysis has not been studied to the same extent.

In this paper we propose a kernel smoothing approach combined with influence function based estimators (Tsiatis, 2007; Newey, 1994; Bickel et al., 1993; Ichimura and Newey, 2015) to deal with continuous treatment for causal mediation analysis. We propose an estimator that, under mild regularity conditions, is consistent, asymptotically linear and well as asymptotically normal. Our work aims to extend the results for continuous treatment ACE for the case of mediation analysis involving continuous treatments in the presence of high-dimensional covariates. (Huber et al., 2020) tackle this problem by weighting the observations by a generalized propensity score that is given as either the conditional density of treatment given (1) the covariates or (2) the covariates and the mediator. The authors estimate the generalized propensity score either parametrically or non-parametrically, and establish asymptotic normality. However, this method inherits the rate of estimation of the generalized propensity score, which can be slow. In contrast, we propose an approach motivated by influence functions and hence obtain many of the desirable properties of influence functions, namely allowing for slower estimation of nuisance parameters, as well as robustness properties. Our work draws heavily from the existing causal mediation literature discussing the identification and estimation of such effects (Pearl, 2001; Imai et al., 2010; Tchetgen Tchetgen and Shpitser, 2012). Additionally, we utilize the double machine learning paradigm from (Chernozhukov et al., 2018).

The rest of the paper is organized as follows. Section 2 introduces the formal mediation framework, describes the identifying assumptions, and discusses an influence function based estimator for binary treatments. Section 3 extends the influence function-based approach to continuous treatment settings, describing the sample-splitting procedure and smoothing procedure. In Section 4, we provide our main result, along with the requisite regularity conditions.

2 Mediation Analysis

Let A be the continuous treatment variable taking values in $\mathcal{A}$, Y be the outcome variable with values in $\mathcal{Y}$, and M be a mediator variable with values in $\mathcal{M}$, which relays parts of the causal effect of A to Y. Also, let X denote observed pre-treatment covariates in the system taking values in $\mathcal{X}$. In order to desribe the causal effect of the treatment on the outcome, we use the potential outcome framework (Pearl, 2001). Let $Y^{(A=a)}$ be the potential outcome variable representing the outcome had (contrary to the fact) the treatment is set to value a. Suppose we are interested in changing the value of the treatment from a to a'. A popular way to measure the causal effect of this change of treatment is to use the average causal effect (ACE), which captures the difference in the expected value of the counterfactual outcome variables, that is

$$ACE = \mathbb{E}[Y^{(a)} - Y^{(a')}],$$

where $\mathbb{E}[\cdot]$ denotes the population-level expectation operator.

The total average causal effect of the treatment on the outcome Y can be partitioned into the part that is mediated by the variable M, and the rest of the causal effect. To formally define this partitioning, let $Y^{(a,m)}$ denote the potential outcome variable corresponding to the outcome had the treatment is set to value a and the mediator is set to value m, and $M^{(a)}$ denote the mediator variable had the treatment is set to value a. Pearl (2001) proposed the following partitioning of the average causal effect into the natural direct and indirect effects:

$$\begin{aligned} ACE &= \overbrace{\mathbb{E}[Y^{(a)} - Y^{(a')}]}^{\text{total effect}} \\ &= \mathbb{E}[Y^{(a,M^{(a)})} - Y^{(a',M^{(a')})}] \\ &= \underbrace{\mathbb{E}[Y^{(a,M^{(a)})} - Y^{(a,M^{(a')})}]}_{\text{natural indirect effect}} + \underbrace{\mathbb{E}[Y^{(a,M^{(a')})} - Y^{(a',M^{(a')})}]}_{\text{natural direct effect}}. \end{aligned} \tag{1}$$

In words, the quantities natural direct effect (NDE) and natural indirect effect (NIE) can be described as follows. NDE captures the change in the expectation of the outcome if the value of the treatment variable is switched between the two arms of the experiment, while the mediator behaves as if the treatment had not changed. NIE captures the change in the expectation of the outcome if the value of the treatment variable is fixed, while the mediator behaves as if the treatment had been switched between the two arms of the experiment. In the following subsection, we discuss estimating NDE and NIE from observational data.

2.1 Estimating Natural Direct and Indirect Effects

In order to estimate the natural direct and indirect effects, from the partitioning in display (1) it is clear that it suffices to focus on estimating quantities of the form

$$\psi_0(a, a') = \mathbb{E}[Y^{(a,M^{(a')})}],$$

for $a, a' \in \mathcal{A}$. Suppose i.i.d. data from distribution P on variables $O = \{A, X, M, Y\}$ is given. In general, the estimand ψ_0 is not identified from observational data and identification assumptions are needed to relate the distribution of the observational data to that of counterfactual variables. We required the following assumptions for the identification

Assumption 1 (Identification Assumptions)

- **Consistency.** *For all $a \in \mathcal{A}$ and $m \in \mathcal{M}$,*

$$\begin{aligned} Y^{(a,m)} &= Y \quad \textit{almost surely if } A = a \textit{ and } M = m, \\ M^{(a)} &= M \quad \textit{almost surely if } A = a. \end{aligned}$$

- **Sequential Exchangeability.** *For all $a, a' \in \mathcal{A}$,*

$$\begin{aligned} (Y^{(a,m)}, M^{(a')}) &\perp A \mid X, \\ Y^{(a,m)} &\perp M \mid A = a', X. \end{aligned}$$

- **Positivity.** *For all* $a \in \mathcal{A}$ *and* $m \in \mathcal{M}$*,*

$$f_{M|A,X}(m|A,X) > 0 \quad \textit{almost surely,}$$
$$f_{A|X}(a|X) > 0 \quad \textit{almost surely,}$$

where $f_{M|A,X}$ *and* $f_{A|X}$ *are the conditional density of* M *given* A *and* X*, and the conditional density of* A *given* X*, respectively.*

Under Assumption 1, the estimand $\psi_0(a, a')$ can be identified using the following expression called the mediation functional, which was originally proposed in (Pearl, 2001).

$$\psi_0(a, a') = \int_{\mathcal{M}} \int_{\mathcal{X}} \mathbb{E}[Y|A = a, M = m, X = x] f_{M|A,X}(m|A = a', X = x) f_X(x) d\mu(m, x), \quad (2)$$

where, f_X is the marginal distribution of X, and μ is a dominating measure for the distribution of (M, X).

Using the expression (2), one can estimate the parameter of interest $\psi_0(a, a')$ by first estimating the nuisance functions $\mathbb{E}[Y|A, M, X]$ and $f_{M|A,X}$, and then using a plug-in estimator to estimate ψ_0 as follows

$$\hat{\psi}_0^{MF}(a, a') = \frac{1}{n} \sum_{i=1}^{n} \int_{\mathcal{M}} \hat{\mathbb{E}}[Y_i|A = a, M = m, X_i] \hat{f}_{M|A,X}(m|A = a', X_i) d\mu(m).$$

Unfortunately, this estimator is sensitive to bias in the estimation of the nuisance functions, that is, misspecification of either of the nuisance functions induces bias in the estimation of the parameter of interest.

As an alternative approach, for the case of binary treatments, i.e., $\mathcal{A} = \{0, 1\}$, Tchetgen Tchetgen and Shpitser (2012) developed a general semiparametric framework for obtaining inferences about the parameter ψ in the nonparametric model with unrestricted observed data model. The authors derived the efficient influence function for $\psi_0(a, a')$

$$\begin{aligned} IF_{\psi_0}(O) = &\frac{I(A = a) f_{M|A,X}(M|A = a', X)}{f_{A|X}(a|X) f_{M|A,X}(M|A = a, X)} \{Y - \mathbb{E}[Y|A = a, M, X]\} \\ &+ \frac{I(A = a')}{f_{A|X}(a'|X)} \{\mathbb{E}[Y|A = a, M, X] - \eta(a, a', X)\} + \eta(a, a', X) - \psi_0(a, a'), \end{aligned} \quad (3)$$

where $a, a' \in \{0, 1\}$, $I(\cdot)$ denotes the indicator function, and

$$\eta(a, a', X) = \int_{\mathcal{M}} \mathbb{E}[Y|A = a, M = m, X] f_{M|A,X}(m|A = a', X) d\mu(m).$$

Note that IF_{ψ_0} is a function of three nuisance functions $\mathbb{E}[Y|A, M, X]$, $f_{M|A,X}$, and $f_{A|X}$. Tchetgen Tchetgen and Shpitser (2012) showed that the estimator based on this influence function has the triple robustness property. That is, it will be consistent even if the model for one (but not more than one) nuisance function is misspecified. Formally, let

- $\mathfrak{M}_{ym}$ be the sub-model in which the model for $\mathbb{E}[Y|A, M, X]$ and $f_{M|A,X}$ are correctly specified.
- $\mathfrak{M}_{ya}$ be the sub-model in which the model for $\mathbb{E}[Y|A, M, X]$ and $f_{A|X}$ are correctly specified.
- $\mathfrak{M}_{ma}$ be the sub-model in which the model for $f_{M|A,X}$ and $f_{M|A,X}$ are correctly specified.

Then the estimator for ψ_0 based on the influence function IF_{ψ_0} is consistent in the submodel $\mathfrak{M}_{ym} \cup \mathfrak{M}_{ya} \cup \mathfrak{M}_{ma}$.

Our goal is to extend the theory of Tchetgen Tchetgen and Shpitser (2012) to the case of continuous treatment variables. This is the objective in the following section.

3 Continuous Mediation Analysis

In this section we show that identification results closely related to the results of Tchetgen Tchetgen and Shpitser (2012) can be established for the case of continuous treatments. The weights in the moment function (3) contain indicator functions which is a problematic feature when dealing with continuous treatments. We modify the weights by utilizing kernel smoothing technique, in which the data with treatment value in a neighborhood defined by bandwidth parameter h is used for weighting each point.

Let d_A denote the dimension of the treatment variable, and let

$$K_h(a) := \frac{1}{h^{d_A}} \prod_{j=1}^{d_A} k\Big(\frac{a_j}{h}\Big),$$

where $k(\cdot)$ is a kernel function, and h denotes the bandwidth parameter. We propose the following moment function for estimating $\psi_0(a, a')$.

$$\begin{aligned} m(O;\alpha,\lambda,\gamma,\psi(a,a')) = {} & K_h(A-a)\lambda(a,X)\frac{\alpha(a',M,X)}{\alpha(a,M,X)}\{Y-\gamma(a,M,X)\} \\ & + K_h(A-a')\lambda(a',X)\{\gamma(a,M,X)-\eta(a,a',X)\} + \eta(a,a',X) - \psi(a,a'), \end{aligned} \tag{4}$$

where $\lambda(a, X) := 1/f_{A|X}(a|X)$, $\alpha(a, M, X) := f_{M|A,X}(M|a, X)$, and $\gamma(a, M, X) := \mathbb{E}[Y|A = a, M, X]$ are the nuisance functions.

We require the kernel function to satisfy the following conditions.

Assumption 2 (Kernel & Bandwidth Assumptions) *the kernel function $k(\cdot)$ satisfies*

- $\int k(u)du = 1$

- $\int uk(u)du = 0$
- $0 < \int u^2 k(u)du < \infty$
- *Bounded Differntiable:* $\left|\frac{dk(u)}{du}\right| < C|u|^{-v}$
- $\int K_h^2(u)du < \infty$

Additionally, the kernel bandwidth is assumed to be a function of the sample size n which satisfies $h \to 0$, $nh^{d_A} \to \infty$ and $nh^{d_A+4} \to C_h$, for a constant C_h, as n goes to infinity.

Note that in the moment function (4), the nuisance parameters are not functions of the parameter of interest ψ. Therefore, it gives us the important property that we do not need to estimate the entire law and having estimators for nuisance functions suffices for obtaining an estimator for the parameter of interest. That is, given estimations of nuisance functions $\hat{\alpha}, \hat{\lambda}, \hat{\gamma}$, the parameter of interest can be estimated as $\hat{\psi}$ by solving the following estimating equation

$$\mathbb{E}[m(O; \hat{\alpha}, \hat{\lambda}, \hat{\gamma}, \hat{\psi}(a, a'))] = 0.$$

3.1 Estimation Procedure

We use cross-fitting estimation approach of Chernozhukov et al. (2018) for separating the estimation of the nuisance functions from the parameter of interest. This approaches provides us with the benefit that weaker smoothness requirements are needed for the nuisance functions. In the cross-fitting approach, we partition the samples into L equal size parts $\{I_1, ..., I_\ell\}$. For $\ell \in \{1, ..., L\}$, we estimate the nuisance functions $\hat{\alpha}_\ell, \hat{\lambda}_\ell, \hat{\gamma}_\ell$ on data from all parts but I_ℓ. For all ℓ, let $\hat{\psi}_\ell$ be the estimation of ψ_0 obtained by solving

$$\frac{1}{|I_\ell|} \sum_{i \in I_\ell} m(O_i; \hat{\alpha}_\ell, \hat{\lambda}_\ell, \hat{\gamma}_\ell, \hat{\psi}_\ell(a, a')) = 0.$$

Our final estimator of ψ_0 is obtained by

$$\hat{\psi}^{TR}(a, a') = \frac{1}{L} \sum_{\ell=1}^{L} \hat{\psi}_\ell(a, a'). \tag{5}$$

We next show that the proposed estimator preserves the triple robustness property of the approach of (Tchetgen Tchetgen and Shpitser, 2012) when the treatments are continuous.

Proposition 1 *The estimator $\hat{\psi}^{TR}(a, a')$ is consistent in the submodel $\mathfrak{M}_{ym} \cup \mathfrak{M}_{ya} \cup \mathfrak{M}_{ma}$.*

4 Asymptotic Analysis

In this section, we study the asymptotic properties of the cross-fitting estimator in display (5), and provide pointwise results. We require the following regularity conditions for the results.

Assumption 3 (Regularity Conditions)

1. *For all M and X, the functions $f(a \mid M, X)$ as a function of a is C^2, and the function and its first and second derivative are bounded.*
2. *The ground truth nuisance functions α, λ, γ and their estimations $\hat{\alpha}, \hat{\lambda}, \hat{\gamma}$ are bounded. Additionally, α and its estimation $\hat{\alpha}$ is bounded away from zero.*
3. $\|([\frac{\partial^2}{\partial a^2} f_{M|A,X}(M|a,X) f_{A|X}(a|X)])\gamma(a,M,X)\|_\infty < \infty$
4. $\|\lambda(a) \left(\frac{\partial^2}{\partial a^2} [f_{M|A,X}(M \mid a, X) f_{A|X}(a \mid X)] \right) \|_\infty < \infty$
5. $\| \frac{\partial^2}{\partial a^2} \left(f_{M|A,X}(M \mid a, X) f_{A|X}(a \mid X) \right) \|_\infty < \infty$
6. $\| \frac{\partial^2}{\partial a^2} \left(\mathbb{E}[Y \mid M, a, X] f_{M|A,X}(M \mid a, X) f_{A|X}(a \mid X) \right) \|_\infty < \infty$

In addition to the regularity conditions, we require the following conditions regarding the convergence of the estimators of the nuisance functions.

Assumption 4 (Consistency)
For any values a and a', the estimators $\frac{\hat{\alpha}(a',M,X)}{\hat{\alpha}(a,M,X)}, \hat{\lambda}(a, X), \hat{\gamma}(a, M, X)$, and $\hat{\eta}(a, a', X)$ are consistent, that is,

i $\int \left(\hat{\lambda}(a, x) - \lambda(a, x) \right)^2 f_X(x) dx \xrightarrow{P} 0$

ii $\int_{\mathcal{X}} \int_{\mathcal{M}} \left(\frac{\hat{\alpha}(a',m,x)}{\alpha(a,\hat{m},x)} - \frac{\alpha(a',m,x)}{\alpha(a,m,x)} \right)^2 f_{M,X}(m, x) dm dx \xrightarrow{P} 0$

iii $\int_{\mathcal{X}} \int_{\mathcal{M}} (\hat{\gamma}(a, m, x) - \gamma(a, m, x))^2 f_{M,X}(m, x) dm dx \xrightarrow{P} 0$

iv $\int_{\mathcal{X}} \left(\hat{\eta}(a, a, x) - \eta(a, a', x)^2 \right) f_X(x) dx \xrightarrow{P} 0$

Assumption 5 (Nuisance Convergence Rates)
For any values a and a', the estimators $\frac{\hat{\alpha}(a',M,X)}{\hat{\alpha}(a,M,X)}, \hat{\lambda}(a, X), \hat{\gamma}(a, M, X)$, and $\hat{\eta}(a, a', X)$ are rate doubly robust, that is,

i $\sqrt{nh^{d_A}} \left(\int_{\mathcal{X}} \int_{\mathcal{M}} \left(\left[\frac{\hat{\alpha}(a',m,x)}{\hat{\alpha}(a,m,x)} - \frac{\alpha(a',m,x)}{\alpha(a,m,x)} \right] \right)^2 f_{M,X}(m, x) dm dx \right)^{\frac{1}{2}} \left(\int_{\mathcal{X}} \int_{\mathcal{M}} (\hat{\gamma}(a, m, x) - \gamma(a, m, x))^2 f_{M,X}(m, x) dm dx \right)^{\frac{1}{2}} \xrightarrow{P} 0$

ii $\sqrt{nh^{d_A}} \left(\int_{\mathcal{X}} \left(\left[\hat{\lambda}(a, x) - \lambda(a, x) \right] \right)^2 f_X(X) dx \right)^{\frac{1}{2}} \left(\int_{\mathcal{X}} \int_{\mathcal{M}} (\hat{\gamma}(a, m, x) - \gamma(a, m, x))^2 f_{M,X}(m, x) dm dx \right)^{\frac{1}{2}} \xrightarrow{P} 0$

iii $\sqrt{nh^{d_A}}\left(\int_{\mathcal{X}}\left(\left[\hat{\lambda}(a,x)-\lambda(a,x)\right]\right)^2 f_X(X)dx\right)^{\frac{1}{2}}\left(\int_{\mathcal{X}}\left([\hat{\eta}(a,a',x)-\eta(a,a',x)]\right)^2 f_X(x)dx\right)^{\frac{1}{2}} \xrightarrow{P} 0$

Note that we do not require any convergence rates on individual nuisance functions. The rate double-robustness condition implies that our requirements on the convergence rate is on the product of the functions. Therefore, if one of the functions converges at a slow rate, the other functions can compensate for that. This is an desired property when working with non-parametric estimators that usually have slow convergence rates.

We have the following result regarding the convergence of the cross-fitting estimator.

Theorem 1 *Suppose Assumptions 2-4 hold. Then for any value pair $a, a' \in \mathcal{A}$, if var() is bounded, then*

$$\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a,a')-\psi_0(a,a')) = \sqrt{\frac{h^{d_A}}{n}}\sum_{i=1}^{n} m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a')) + o_p(1),$$

and $\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a,a')-\psi_0(a,a')-B(a,a'))$ converges to the Gaussian distribution $\mathcal{N}(0, V(a,a'))$, with

$$\begin{aligned} B(a,a') =& h^2\int u^2k(u)du \\ &\times \mathbb{E}\Bigg\{\frac{f(M\mid A=a',X)}{f(M\mid A=a,X)}\left(\partial_a\gamma(X,M,a)\frac{\partial_a f(a|X,M)}{f(a\mid X)}+\frac{1}{2}\frac{f(a\mid X,M)}{f(a\mid X)}\partial_a^2\gamma(X,M,a)\right) \\ &+\{\mathbb{E}[Y\mid X,M,A=a]-\eta(a,a',X)\}\frac{1}{2}\frac{\partial_a^2 f(a'|X,M)}{f(a'|X)}\Bigg\} + O(h^3), \end{aligned}$$

and

$$V(a,a') = \int k(u)^2du \times E_1 + O(h)$$

where $E_1 = \mathbb{E}\left\{\frac{f(M|a',X)^2 f(a|X,M)}{f(M|a,X)^2 f(a|X)^2} var(Y|X,M,a) + \frac{1}{f(a'|X)} var[E(Y|X,M,a)|X,a']\right\}$.

We next extend the result in Theorem 1 to uniformity over a compact interior of the support of the treatment variable. We require the following stronger version of Assumption 4 for our result.

Assumption 6 *Let $\mathcal{A}_0$ be a compact interior of the support of the treatment variable. We assume that for the supremum of the pair $(a,a') \in \mathcal{A}_0$, the consistency and rate double robustness conditions in Assumption 4 hold. Moreover, the estimators $\frac{\hat{\alpha}(a',\cdot,\cdot)}{\hat{\alpha}(a,\cdot,\cdot)}$, $\hat{\lambda}(a,\cdot)$, $\hat{\gamma}(a,\cdot,\cdot)$, and $\hat{\eta}(a,a',\cdot)$ are Lipschitz continuous in $\mathcal{A}_0$.*

i $\sup_{a',a\in\mathcal{A}_0}\sqrt{nh^{d_A}}\left(\int_{\mathcal{X}\times\mathcal{M}}\left(\left[\frac{\hat{\alpha}(a',m,x)}{\hat{\alpha}(a,m,x)}-\frac{\alpha(a',m,x)}{\alpha(a,m,x)}\right]\right)^2 f_{M,X}(m,x)dmdx\right)^{\frac{1}{2}}\left(\int_{\mathcal{X}\times\mathcal{M}}(\hat{\gamma}(a,m,x)-\gamma(a,m,x))^2 f_{M,X}(m,x)dmdx\right)^{\frac{1}{2}} \xrightarrow{P} 0$

ii $\sup_{a',a\in\mathcal{A}_0}\sqrt{nh^{d_A}}\left(\int_{\mathcal{X}}\left(\left[\hat{\lambda}(a,x)-\lambda(a,x)\right]\right)^2 f_X(X)dx\right)^{\frac{1}{2}}\left(\int_{\mathcal{X}\times\mathcal{M}}(\hat{\gamma}(a,m,x)-\gamma(a,m,x))^2 f_{M,X}(m,x)dmdx\right)^{\frac{1}{2}} \xrightarrow{P} 0$

iii $\sup_{a',a\in\mathcal{A}_0}\sqrt{nh^{d_A}}\left(\int_{\mathcal{X}}\left(\left[\hat{\lambda}(a,x)-\lambda(a,x)\right]\right)^2 f_X(X)dx\right)^{\frac{1}{2}}\left(\int_{\mathcal{X}}([\hat{\eta}(a,a',x)-\eta(a,a',x)])^2 f_X(x)dx\right)^{\frac{1}{2}} \xrightarrow{P} 0$

Theorem 2 *Let the conditions in Theorem 1 and Assumption 6 hold. Then the asymptotically linear representation of in Theorem 1 holds uniformly over* $\mathcal{A}_0$.

5 Conclusion

In this paper we put forward a flexible method to estimate mediation effects in the presence of continuous treatments and high dimensional covariates. We provide results on rate as well as asymptotic normality to allow for the construction of confidence intervals and hypothesis tests. Extending the point-wise results to uniform results and providing optimal rates is left for future work.

Acknowledgment

We are grateful to Prof. Eric Tchetgen Tchetgen for helpful discussions and comments.

References

Ai, C., Linton, O., and Zhang, Z. (2021). Estimation and inference for the counterfactual distribution and quantile functions in continuous treatment models. *Journal of Econometrics*.

Bickel, P. J., Klaassen, C. A., Bickel, P. J., Ritov, Y., Klaassen, J., Wellner, J. A., and Ritov, Y. (1993). *Efficient and adaptive estimation for semiparametric models*, volume 4. Johns Hopkins University Press Baltimore.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Colangelo, K. and Lee, Y.-Y. (2020). Double debiased machine learning nonparametric inference with continuous treatments. *arXiv preprint arXiv:2004.03036*.

Goetgeluk, S., Vansteelandt, S., and Goetghebeur, E. (2008). Estimation of controlled direct effects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):1049–1066.

Hernán, M. A. and Robins, J. M. (2020). Causal inference: what if.

Hill, J. L. (2011). Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 20(1):217–240.

Hirano, K. and Imbens, G. W. (2004). *The Propensity Score with Continuous Treatments*, chapter 7, pages 73–84. John Wiley & Sons, Ltd.

Huber, M., Hsu, Y.-C., Lee, Y.-Y., and Lettry, L. (2020). Direct and indirect effects of continuous treatments based on generalized propensity score weighting. *Journal of Applied Econometrics*, 35(7):814–840.

Ichimura, H. and Newey, W. K. (2015). The influence function of semiparametric estimators.

Imai, K., Keele, L., and Yamamoto, T. (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical science*, pages 51–71.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika*, 87(3):706–710.

Kallus, N. and Zhou, A. (2018). Policy evaluation and optimization with continuous treatments. In *International Conference on Artificial Intelligence and Statistics*, pages 1243–1251. PMLR.

Kennedy, E. H., Ma, Z., McHugh, M. D., and Small, D. S. (2017). Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 79(4):1229.

Kreif, N., Grieve, R., Díaz, I., and Harrison, D. (2015). Evaluation of the effect of a continuous treatment: a machine learning approach with an application to treatment for traumatic brain injury. *Health economics*, 24(9):1213–1228.

Lange, T. and Hansen, J. V. (2011). Direct and indirect effects in a survival context. *Epidemiology*, pages 575–581.

Lange, T., Vansteelandt, S., and Bekaert, M. (2012). A simple unified approach for estimating natural direct and indirect effects. *American journal of epidemiology*, 176(3):190–195.

Neugebauer, R. and van der Laan, M. (2007). Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica: Journal of the Econometric Society*, pages 1349–1382.

Pearl, J. (2001). Direct and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI-01).*

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 95–133. Springer.

Robins, J. M. and Greenland, S. (1992). Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, pages 143–155.

Robins, J. M. and Rotnitzky, A. (1995). Semiparametric efficiency in multivariate regression models with missing data. *Journal of the American Statistical Association*, 90(429):122–129.

Su, L., Ura, T., and Zhang, Y. (2019). Non-separable models with high-dimensional data. *Journal of Econometrics*, 212(2):646–677.

Tchetgen Tchetgen, E. and Shpitser, I. (2012). Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Annals of statistics*, 40(3):1816.

Tsiatis, A. (2007). *Semiparametric theory and missing data.* Springer Science & Business Media.

van der Laan, M. J. and Petersen, M. L. (2008). Direct effect models. *The international journal of biostatistics*, 4(1).

Van Der Laan, M. J. and Robins, J. M. (1998). Locally efficient estimation with current status data and time-dependent covariates. *Journal of the American Statistical Association*, 93(442):693–701.

VanderWeele, T. J. (2009). Marginal structural models for the estimation of direct and indirect effects. *Epidemiology*, pages 18–26.

6 Appendix

Before we start with the proofs, we establish some lemmas that will help us with the proofs in the rest of the appendix.

Lemma 1 *Let $\{X_m\}$ and $\{Y_m\}$ be a sequence of random variables. Then under conditions outlined in Lemma 6.1 in Chernozhukov et al. (2018), $\mathbb{E}[|X_m| \mid Y_m] = o_p(1)$ implies $X_m = o_p(1)$.*

Proof: [Proof of Lemma 1]

By the Conditional Markov Inequality, for any $\epsilon > 0$,

$$p(|X_m| \geq \epsilon \mid Y_m) \leq \frac{\mathbb{E}[|X_m| \mid Y_m]}{\epsilon}$$

By $\mathbb{E}[|X_m| \mid Y_m] = o_p(1)$, there is $p(|X_m| \geq \epsilon \mid Y_m) = o_p(1)$. An application of Lemma 6.1 then yields $p(|X_m| > \epsilon) \to 0$, therefore $X_m = o_p(1)$. □

Lemma 2 *Under Assumption 2, we have*

$$\int_A K_h(A-a)f(A)dA = f(a) + Ch^2 f''(a) + O_p(h^3).$$

Proof: [Proof of Lemma 2]

$$\begin{aligned}
\int_A K_h(A-a)f(A)dA &= \int_u k(u)f(uh+a)du \\
&= \int_u k(u)\{f(a) + uhf'(a) + u^2h^2f''(a) + O_p(u^3h^3)\}du \\
&= f(a) + Ch^2 f''(a) + O_p(h^3),
\end{aligned}$$

where the last equality follows from the assumptions that $\int k(u)du = 1$, $\int k(u)udu = 0$, and $\int k(u)u^2du$ is constant.

What did we decide to do with the term in red?

6.1 Proof for Theorem 1

We follow a similar outline as Colangelo and Lee (2020) and Chernozhukov et al. (2018). The proof for this theorem is split into two parts. The first part establishes that the proposed estimator satisfies

$$\sqrt{\frac{h^{d_A}}{n}} \sum_{\ell=1}^{L} \sum_{i \epsilon I_\ell} \Big\{ m(O_i; \hat{\alpha}_\ell, \hat{\lambda}_\ell, \hat{\gamma}_\ell, \psi_0(a,a')) - m(O_i; \alpha, \lambda, \gamma, \psi_0(a,a')) \Big\} = o_p(1),$$

and the second part establishes that $\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a,a') - \psi_0(a,a') - B(a,a'))$ converges to the Gaussian distribution $\mathcal{N}(0, V(a,a'))$.

Starting with the first part of the proof, note that

$$\begin{aligned}
&\sqrt{nh^{d_A}}\frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\Big\{m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\hat{\psi}_\ell(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\}\\
=&\sqrt{\frac{h^{d_A}}{n}}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\Big\{m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\hat{\psi}_\ell(a,a')) - m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a'))\\
&+ m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\}\\
=&-\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a,a') - \psi_0(a,a'))\\
&+\sqrt{\frac{h^{d_A}}{n}}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\Big\{m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\}.
\end{aligned}$$

Since $\frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in I_\ell} m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\hat{\psi}_\ell(a,a')) = 0$, we have

$$\begin{aligned}
&\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a,a') - \psi_0(a,a'))\\
=&\sqrt{\frac{h^{d_A}}{n}}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\Big\{m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\}\\
&+\sqrt{\frac{h^{d_A}}{n}}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\Big\{m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\}.
\end{aligned}$$

In order to establish an asymptotically linear representation for our proposed estimator, it suffices to to show that for all $1 \le \ell \le L$ we have

$$\sqrt{\frac{h^{d_A}}{n}}\sum_{i\in I_\ell}\Big\{m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\} = o_p(1).$$

Next, we expand $m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))$ into multiple terms and bound each term individually. Note that

$$\begin{aligned}
&m(O_i;\hat{\alpha}_\ell,\hat{\lambda}_\ell,\hat{\gamma}_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\\
&= K_h(A_i - a)\big\{\hat{\lambda}(a,X_i)\frac{\hat{\alpha}(a',M_i,X_i)}{\hat{\alpha}(a,M_i,X_i)}\{Y_i - \hat{\gamma}(a,M_i,X_i)\} - \lambda(a,X_i)\frac{\alpha(a',M_i,X_i)}{\alpha(a,M_i,X_i)}\{Y_i - \gamma(a,M_i,X_i)\}\big\}\\
&\quad + K_h(A_i - a')\big\{\hat{\lambda}(a',X_i)\{\hat{\gamma}(a,M_i,X_i) - \hat{\eta}(a,a',X_i)\} - \lambda(a',X_i)\{\gamma(a,M_i,X_i) - \eta(a,a',X_i)\}\big\}\\
&\quad + \hat{\eta}(a,a',X_i) - \eta(a,a',X_i).
\end{aligned}$$

To make the notations more concise, with slight abuse of notation, for a given a and a', we define $\lambda_a(X_i) := \lambda(a,X_i)$, $R(M_i,X_i) := \frac{\alpha(a',M_i,X_i)}{\alpha(a,M_i,X_i)}$, $\gamma_a(M_i,X_i) := \gamma(a,M_i,X_i)$, and $\eta(X_i) := \eta(a,a',X_i)$.

$$
\begin{aligned}
&m(O_i;\hat\alpha_\ell,\hat\lambda_\ell,\hat\gamma_\ell,\psi_0(a,a')) - m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a')) \\
&= K_h(A_i-a)\big\{\hat\lambda_a(X_i)\hat R(M_i,X_i)\{Y_i-\hat\gamma_a(M_i,X_i)\} - \lambda_a(X_i)R(M_i,X_i)\{Y_i-\gamma_a(M_i,X_i)\}\big\} && (6)\\
&\quad + K_h(A_i-a')\big\{\hat\lambda_{a'}(X_i)\{\hat\gamma_a(M_i,X_i)-\hat\eta(X_i)\} - \lambda_{a'}(X_i)\{\gamma_a(M_i,X_i)-\eta(X_i)\}\big\} && (7)\\
&\quad + \hat\eta(X_i)-\eta(X_i). && \text{(R1)}
\end{aligned}
$$

Terms (6) and (7) are expanded additionally. Expanding term (6), we get

$$
\begin{aligned}
&K_h(A_i-a)\big\{\hat\lambda_a(X_i)\hat R(M_i,X_i)\{Y_i-\hat\gamma_a(M_i,X_i)\} - \lambda_a(X_i)R(M_i,X_i)\{Y_i-\gamma_a(M_i,X_i)\}\big\} \\
&= -K_h(A_i-a)\big(\hat R(M_i,X_i)-R(M_i,X_i)\big)\big(\hat\lambda_a(X_i)-\lambda_a(X_i)\big)\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big) && \text{(CS1)}\\
&\quad + K_h(A_i-a)\big(\hat R(M_i,X_i)-R(M_i,X_i)\big)\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big)\big(Y_i-\gamma_a(M_i,X_i)\big) && \text{(CS2)}\\
&\quad - K_h(A_i-a)\big(\hat R(M_i,X_i)-R(M_i,X_i)\big)\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big)\lambda_a(X_i) && \text{(CS3)}\\
&\quad - K_h(A_i-a)\big(\hat\lambda_a(X_i)-\lambda_a(X_i)\big)\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big)R(M_i,X_i) && \text{(CS4)}\\
&\quad + \Big\{K_h(A_i-a)\big(\hat R(M_i,X_i)-R(M_i,X_i)\big)\lambda_a(X_i)\big(Y_i-\gamma_a(M_i,X_i)\big) \\
&\qquad - \mathbb{E}\big[K_h(A_i-a)\big(\hat R(M_i,X_i)-R(M_i,X_i)\big)\lambda_a(X_i)\big(Y_i-\gamma_a(M_i,X_i)\big)\big]\Big\} && \text{(E1)}\\
&\quad + \mathbb{E}\big[K_h(A_i-a)\big(\hat R(M_i,X_i)-R(M_i,X_i)\big)\lambda_a(X_i)\big(Y_i-\gamma_a(M_i,X_i)\big)\big] && \text{(TR1)}\\
&\quad + \Big\{K_h(A_i-a)\big(\hat\lambda_a(X_i)-\lambda_a(X_i)\big)R(M_i,X_i)\big(Y_i-\gamma_a(M_i,X_i)\big) \\
&\qquad - \mathbb{E}\big[K_h(A_i-a)\big(\hat\lambda_a(X_i)-\lambda_a(X_i)\big)R(M_i,X_i)\big(Y_i-\gamma_a(M_i,X_i)\big)\big]\Big\} && \text{(E2)}\\
&\quad + \mathbb{E}\big[K_h(A_i-a)\big(\hat\lambda_a(X_i)-\lambda_a(X_i)\big)R(M_i,X_i)\big(Y_i-\gamma_a(M_i,X_i)\big)\big] && \text{(TR2)}\\
&\quad - K_h(A_i-a)\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big)\lambda_a(X_i)R(M_i,X_i). && \text{(R2)}
\end{aligned}
$$

For term (7), note that

$$
\begin{aligned}
&K_h(A_i-a')\big\{\hat\lambda_{a'}(X_i)\{\hat\gamma_a(M_i,X_i)-\hat\eta(X_i)\} - \lambda_{a'}(X_i)\{\gamma_a(M_i,X_i)-\eta(X_i)\}\big\} \\
&= K_h(A_i-a')\big(\hat\lambda_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big) && \text{(CS5)}\\
&\quad - K_h(A_i-a')\big(\hat\lambda_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big(\hat\eta(X_i)-\eta(X_i)\big) && \text{(CS6)}\\
&\quad + K_h(A_i-a')\big(\hat\lambda_{a'}(X_i)-\lambda_{a'}(X_i)\big)\gamma_a(M_i,X_i) && \text{(R3)}\\
&\quad + K_h(A_i-a')\big(\hat\gamma_a(M_i,X_i)-\gamma_a(M_i,X_i)\big)\lambda_{a'}(X_i) && \text{(R4)}\\
&\quad - K_h(A_i-a')\big(\hat\lambda_{a'}(X_i)-\lambda_{a'}(X_i)\big)\eta(X_i) && \text{(R5)}\\
&\quad - K_h(A_i-a')\big(\hat\eta(X_i)-\eta(X_i)\big)\lambda_{a'}(X_i). && \text{(R6)}
\end{aligned}
$$

Next, we group terms (R1)-(R6) as follows. We pair (R1) with (R6), (R2) with (R4), and (R3) with (R5). Note that every expectation introduced here is only over O_i, conditional on O_ℓ^c, i.e., $\mathbb{E}(\cdot|O_\ell^c)$, and hence all the terms are random variables. For (R1)+(R6) we have

$$
\begin{aligned}
&(R1) + (R6) \\
&= \big(\hat{\eta}(X_i) - \eta(X_i)\big) - K_h(A_i - a')\big(\hat{\eta}(X_i) - \eta(X_i)\big)\lambda_{a'}(X_i) \\
&= \big(\hat{\eta}(X_i) - \eta(X_i)\big) - \mathbb{E}\big[\hat{\eta}(X_i) - \eta(X_i)\big] && \text{(E3)} \\
&\quad - \Big\{K_h(A_i - a')\big(\hat{\eta}(X_i) - \eta(X_i)\big)\lambda_{a'}(X_i) - \mathbb{E}\big[K_h(A_i - a')\big(\hat{\eta}(X_i) - \eta(X_i)\big)\lambda_{a'}(X_i)\big]\Big\} && \text{(E4)} \\
&\quad + \mathbb{E}\big[\big(\hat{\eta}(X_i) - \eta(X_i)\big)\big(1 - K_h(A_i - a')\lambda_{a'}(X_i)\big)\big]. && \text{(TR3)}
\end{aligned}
$$

For (R2)+(R4) we have

$$
\begin{aligned}
&(R2) + (R4) \\
&= -K_h(A_i - a)\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_a(X_i)R(M_i, X_i) + \\
&\qquad\qquad K_h(A_i - a')\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i) \\
&= -\Big\{K_h(A_i - a)\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_a(X_i)R(M_i, X_i) \\
&\qquad - \mathbb{E}\big[K_h(A_i - a)\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_a(X_i)R(M_i, X_i)\big]\Big\} && \text{(E5)} \\
&\quad + \Big\{K_h(A_i - a')\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i) \\
&\qquad - \mathbb{E}\big[K_h(A_i - a')\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i)\big]\Big\} && \text{(E6)} \\
&\quad + \mathbb{E}\big[\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\big\{K_h(A_i - a')\lambda_{a'}(X_i) - K_h(A_i - a)\lambda_a(X_i)R(M_i, X_i)\big\}\big]. && \text{(TR4)}
\end{aligned}
$$

For (R3)+(R5) we have

$$
\begin{aligned}
&(R3) + (R5) \\
&= K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\gamma_a(M_i, X_i) - K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\eta(X_i) \\
&= \Big\{K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\gamma_a(M_i, X_i) \\
&\qquad - \mathbb{E}\big[K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\gamma_a(M_i, X_i)\big]\Big\} && \text{(E7)} \\
&\quad - \Big\{K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\eta(X_i) \\
&\qquad - \mathbb{E}\big[K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\eta(X_i)\big]\Big\} && \text{(E8)} \\
&\quad + \mathbb{E}\big[K_h(A_i - a')\big(\hat{\lambda}_{a'}(X_i) - \lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i, X_i) - \eta(X_i)\big\}\big]. && \text{(TR5)}
\end{aligned}
$$

And so, to prove $\sqrt{\frac{h^{d_A}}{n}} \sum_{i \epsilon I_\ell} \left\{ m(O_i; \hat{\alpha}_\ell, \hat{\lambda}_\ell, \hat{\gamma}_\ell, \psi_0(a,a')) - m(O_i; \alpha, \lambda, \gamma, \psi_0(a,a')) \right\} = o_p(1)$, we need $\sqrt{nh^{d_A}}$ times each of the terms (CS1) - (CS6), (E1) - (E8) and (TR1) - (TR5) to be $o_p(1)$.

6.1.1 Proof for Terms (CS1)-(CS6)

All of these terms contain the product of two or more errors and can be treated similarly. We provide a detailed proof for (CS2), and a similar method can be followed for the rest of the terms.

Starting with (CS2), write $\Delta_{i\ell} = K_h(A_i - a)\big[\hat{R}(M_i, X_i) - R(M_i, X_i)\big]\left[\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\right]\big[Y_i - \gamma_a(M_i, X_i)\big]$. Following lemma 1, it suffices to bound $\mathbb{E}\left[|\sqrt{\frac{h^{d_A}}{n}} \sum_{i \in I_\ell} \Delta_{i\ell}| \mid O_\ell^C \right]$ as $o_p(1)$ in order to show that $\sqrt{\frac{h^{d_A}}{n}} \sum_{i \in I_\ell} \Delta_{i\ell} = o_p(1)$.

First, from the triangle inequality, $\mathbb{E}\left[\left| \sqrt{\frac{h^{d_A}}{n}} \sum_{i \in I_\ell} \Delta_{i\ell} \right| \mid O_\ell^C \right] \leq \sqrt{nh^{d_A}} \sum_{i \in I_\ell} \mathbb{E}\left[|\Delta_{i\ell}| \mid O_\ell^C \right]$, and so it suffices to bound $\sqrt{nh^{d_A}} \mathbb{E}\left[\left|\Delta_{i\ell}\right| \middle| O_\ell^c \right]$

$$
\begin{aligned}
&\sqrt{nh^{d_A}} \mathbb{E}\left[\left|\Delta_{i\ell}\right| \middle| O_\ell^c \right] \\
=& \sqrt{nh^{d_A}} \int_{\mathcal{O}} \left| K_h(A_i - a)\big[\hat{R}(M_i, X_i) - R(M_i, X_i)\big]\left[\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\right]\left[Y_i - \gamma_a(M_i, X_i)\right] \right| \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad f(Y_i, A_i, M_i, X_i) dO_i \\
=& \sqrt{nh^{d_A}} \int_{\mathcal{O}} \left| k(u)\big[\hat{R}(M_i, X_i) - R(M_i, X_i)\big]\left[\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\right]\left[Y_i - \gamma_a(M_i, X_i)\right] \right| \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad f(Y_i, uh + a, M_i, X_i) du dY_i dM_i dX_i \\
=& \sqrt{nh^{d_A}} \int_{\mathcal{M} \times \mathcal{X}} \left\{ \int_{\mathcal{U}} \left| k(u) f(uh + a | M_i, X_i) \left\{ \int_{\mathcal{Y}} \left[Y_i - \gamma_a(M_i, X_i)\right] f(Y_i | uh + a, M_i, X_i) dY_i \right\} \right| du \right\} \\
&\qquad \left| \big[\hat{R}(M_i, X_i) - R(M_i, X_i)\big]\left[\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\right] f(M_i, X_i) \right| dM_i dX_i \\
=& \sqrt{nh^{d_A}} \int_{\mathcal{M} \times \mathcal{X}} \left\{ \int_{\mathcal{U}} \left| \left[\gamma_{uh+a}(M_i, X_i) - \gamma_a(M_i, X_i)\right] \right| k(u) f(uh + a | M_i, X_i) du \right\} \\
&\qquad \left| \big[\hat{R}(M_i, X_i) - R(M_i, X_i)\big]\left[\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\right] \right| f(M_i, X_i) dM_i dX_i
\end{aligned}
$$

Next, by Assumption 3, Lemma 2 and Assumption 5 (i), the term can be shown to be bounded.

$$
\begin{aligned}
=&O_p(\sqrt{nh^{d_A}})\int_{\mathcal{M}\times\mathcal{X}}\left\{\int_{\mathcal{U}}k(u)f(a|M_i,X_i)+k(u)u^2O_p(h^2)du\right\} \\
&\qquad\left|\left[\hat{R}(M_i,X_i)-R(M_i,X_i)\right]\left[\hat{\gamma}_a(M_i,X_i)-\gamma_a(M_i,X_i)\right]\right|f(M_i,X_i)dM_idX_i \\
\leq&O_p\Bigg(\Bigg\{\sqrt{nh^{d_A}}\int_{\mathcal{M}\times\mathcal{X}}\left[\hat{R}(M_i,X_i)-R(M_i,X_i)\right]^2f(M_i,X_i)dM_idX_i \\
&\qquad\int_{\mathcal{M}\times\mathcal{X}}\left[\hat{\gamma}_a(M_i,X_i)-\gamma_a(M_i,X_i)\right]^2f(M_i,X_i)dM_idX_i\Bigg\}^{1/2}\Bigg) \\
=&o_p(1).
\end{aligned}
$$

6.1.2 Proof for Terms (E1)-(E8)

Terms (E1)-(E8) are normalized terms of the form of a bias times a bounded quantity; they can all be treated similarly. We only provide the proof of the convergence in probability to zero for the term (E2).

Assumptions for E1-8:

- Y_i's conditional mean $\gamma_a(m,x)$, and conditional variance $var(Y_i|a,m,x)$ are bounded over $a\in\mathcal{A}_0$ for any $m\in\mathcal{M}$ and $x\in\mathcal{X}$
- $R(M_i,X_i)$, $\lambda_a(X_i)$, $\gamma_a(M_i,X_i)$, and $\eta(X_i)$ are bounded over $a',a\in\mathcal{A}_0$, $(M_i,X_i)\in\mathcal{M}\times\mathcal{X}$
- Assumption 4: for all $a',a\in\mathcal{A}_0$,
    1. $\int_{\mathcal{X}}\left[\hat{\lambda}_a(X_i)-\lambda_a(X_i)\right]^2f(X_i)dX_i\xrightarrow{P}0$
    2. $\int_{\mathcal{M}\times\mathcal{X}}\left[\hat{R}(M_i,X_i)-R(M_i,X_i)\right]^2f(M_i,X_i)dM_idX_i\xrightarrow{P}0$
    3. $\int_{\mathcal{M}\times\mathcal{X}}\left[\hat{\gamma}_a(M_i,X_i)-\gamma_a(M_i,X_i)\right]^2f(M_i,X_i)dM_idX_i\xrightarrow{P}0$
    4. $\int_{\mathcal{X}}\left[\hat{\eta}(X_i)-\eta(X_i)\right]^2f(X_i)dX_i\xrightarrow{P}0$

Proof for (E2). There are $\lambda_a(X_i)=\lambda(a,X_i)$, $R(M_i,X_i)=\frac{\alpha(a',M_i,X_i)}{\alpha(a,M_i,X_i)}$, and $\gamma_a(M_i,X_i)=\gamma(a,M_i,X_i)$.

Set

$$
\begin{aligned}
\hat{\Delta}_{i\ell}=&K_h(A_i-a)\left[\hat{\lambda}_a(X_i)-\lambda_a(X_i)\right]R(M_i,X_i)\left[Y_i-\gamma_a(M_i,X_i)\right]- \\
&\mathbb{E}\Bigg\{K_h(A_i-a)\left[\hat{\lambda}_a(X_i)-\lambda_a(X_i)\right]R(M_i,X_i)\left[Y_i-\gamma_a(M_i,X_i)\right]\Bigg\}.
\end{aligned}
$$

By construction and independence of O_ℓ^c and O_i, $i \in I_\ell$, $\mathbb{E}(\hat{\Delta}_{i\ell}|O_\ell^c) = 0$, and $\mathbb{E}(\hat{\Delta}_{i\ell}\hat{\Delta}_{j\ell}|O_\ell^c) = 0$ for $i, j \in I_\ell$ and all $a', a \in \mathcal{A}_0$.

$$
\begin{aligned}
& h^{d_A}\mathbb{E}(\hat{\Delta}_{i\ell}^2|O_\ell^c) \\
=& h^{d_A}\int_{\mathcal{O}} K_h^2(A_i - a)\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]^2 R_i^2(a', a)\left[Y_i - \gamma_a(M_i, X_i)\right]^2 f(Y_i, A_i, M_i, X_i)dO_i \\
=& \int_{\mathcal{O}} k^2(u)\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]^2 R_i^2(a', a)\left[Y_i - \gamma_a(M_i, X_i)\right]^2 f(Y_i, uh + a, M_i, X_i)dudY_idM_idX_i \\
=& \int_{\mathcal{M}\times\mathcal{X}}\int_{\mathcal{U}} k^2(u)f(uh + a|M_i, X_i)\left\{\int_{\mathcal{Y}}\left[Y_i - \gamma_a(M_i, X_i)\right]^2 f(Y_i|uh + a, M_i, X_i)dY_i\right\}du \\
& \qquad \left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]^2 R_i^2(a', a)f(M_i, X_i)dM_idX_i \\
\leq& O_p\left(\int_{\mathcal{U}} k^2(u)du\int_{\mathcal{M}\times\mathcal{X}}\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]^2 R_i^2(a', a)f(M_i, X_i)dM_idX_i\right) \\
=& o_p(1)
\end{aligned}
$$

where the $O_p(\cdot)$ statement comes from the boundedness of Y_i's conditional mean $\gamma_a(M_i, X_i)$ and conditional variance $var(Y_i|a, m, x)$ over $a \in \mathcal{A}_0$ for any $m \in \mathcal{M}$ and $x \in \mathcal{X}$, and that

$$
\begin{aligned}
& \int_{\mathcal{Y}}\left[Y_i - \gamma_a(M_i, X_i)\right]^2 f(Y_i|uh + a, M_i, X_i)dY_i \\
=& \int_{\mathcal{Y}}\left[Y_i^2 + \gamma_a^2(M_i, X_i) - 2\gamma_a(M_i, X_i)Y_i\right] f(Y_i|uh + a, M_i, X_i)dY_i \\
=& \mathbb{E}[Y_i^2|uh + a, M_i, X_i] + \gamma_a^2(M_i, X_i) - 2\gamma_a(M_i, X_i)\int_{\mathcal{Y}} Y_i f(Y_i|uh + a, M_i, X_i)dY_i \\
=& \mathbb{E}[Y_i^2|uh + a, M_i, X_i] + \gamma_a^2(M_i, X_i) - 2\gamma_a(M_i, X_i) - 2\gamma_a(M_i, X_i)\gamma_{uh+a}(M_i, X_i) \\
=& O_p(1).
\end{aligned}
$$

The $o_p(1)$ rate is from assuming $R(M_i, X_i)$ is bounded over $a', a \in \mathcal{A}_0$ and assumption 4(i) that

$$
\int_{\mathcal{X}}\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]^2 f(X_i)dX_i \xrightarrow{P} 0.
$$

Then

$$
\mathbb{E}\left[\left(\sqrt{h^{d_A}/n}\sum_{l=1}^{L}\sum_{i\in I_\ell}\hat{\Delta}_{i\ell}\right)^2\middle|O_\ell^c\right] = h^{d_A}/n\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\mathbb{E}(\hat{\Delta}_{i\ell}^2|O_\ell^c) = h^{d_A}\mathbb{E}(\hat{\Delta}_{i\ell}^2|O_\ell^c) = o_p(1).
$$

This leads to $\sqrt{h^{d_A}/n}\sum_{l=1}^{L}\sum_{i\in I_\ell}\hat{\Delta}_{i\ell} \xrightarrow{P} 0$, i.e. (E2) being $o_p(1/\sqrt{nh^{d_A}})$.

6.1.3 Proof for Terms (TR1)-(TR5)

The proofs of the convergence in probability to zero for the terms (TR1)-(TR5) require extra considerations, and we prove them on a case by case basis below. Since these terms are already in expectation form, we don't need to invoke Lemma 1 here.

Proof for Terms TR1 and TR2

Terms (TR1) and (TR2) are similar; we only provide the proof of the convergence in probability to zero for the term (TR2).

To bound TR2, first let $\hat{\Delta}_{i\ell} = K_h(A_i - a)\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i)\left[Y_i - \gamma_a(M_i, X_i)\right]$. Bounding (TR2) amounts to showing $\sqrt{nh^{d_A}}\mathbb{E}(\hat{\Delta}_{i\ell}) = o_p(1)$.

$$
\begin{aligned}
&\sqrt{nh^{d_A}}\mathbb{E}\left(\Delta_{i\ell}\middle|O_\ell^c\right)\\
=&\sqrt{nh^{d_A}}\mathbb{E}\left\{K_h(A_i - a)\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i)\left[Y_i - \gamma_a(M_i, X_i)\right]\middle|O_\ell^c\right\}\\
=&\sqrt{nh^{d_A}}\int_{\mathcal{O}} K_h(A_i - a)\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i)\left[Y_i - \gamma_a(M_i, X_i)\right] f(Y_i, A_i, M_i, X_i)dO_i\\
=&\sqrt{nh^{d_A}}\int_{\mathcal{M}\times\mathcal{X}\times\mathcal{Y}}\left[\int_{\mathcal{A}} K_h(A_i - a)f(A_i \mid Y_i, M_i, X_i)\right]\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]\\
&\qquad R(M_i, X_i)\left[Y_i - \gamma_a(M_i, X_i)\right] f(Y_i, M_i, X_i)dY_i dM_i dX_i
\end{aligned}
$$

Applying Lemma 2 under Assumption 3.2

$$
\begin{aligned}
=&\sqrt{nh^{d_A}}\int_{\mathcal{M}\times\mathcal{X}\times\mathcal{Y}}\left[f(a \mid Y_i, M_i, X_i) + C^2h^2 f''(a \mid M_i, X_i) + O_p(h^3)\right]\\
&\qquad\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i)\left[Y_i - \gamma_a(M_i, X_i)\right] f(Y_i, M_i, X_i)dY_i dM_i dX_i\\
\overset{(a)}{=}&\sqrt{nh^{d_A}}\int_{\mathcal{M}\times\mathcal{X}\times\mathcal{Y}}\left[C^2h^2 f''(a \mid M_i, X_i) + O_p(h^3)\right]\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]\\
&\qquad R(M_i, X_i)\left[Y_i - \gamma_a(M_i, X_i)\right] f(Y_i, M_i, X_i)dY_i dM_i dX_i\\
=&\sqrt{nh^{d_A}}\int_{\mathcal{M}\times\mathcal{X}}\left[C^2h^2 f''(a \mid M_i, X_i) + O_p(h^3)\right]\left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right]\\
&\qquad R(M_i, X_i)\left[\int_{\mathcal{Y}} Y_i f(Y_i \mid M_i, X_i)dY_i - \gamma_a(M_i, X_i)\right] f(M_i, X_i)dM_i dX_i\\
=&o_p(1)
\end{aligned}
$$

The last equality follows from Assumption 2 ($nh^{d_A+4} \to C_h$, $h \to 0$), the boundedness of $f''(a \mid M_i, X_i)$ and $R(M_i, X_i)$, the consistency of $\hat{\lambda}_a(X_i)$ in Assumption 4.1, and the boundedness of $\int_{\mathcal{Y}} Y_i f(Y_i \mid M_i, X_i) dY_i$ from that

$$\begin{aligned}
&\int_{\mathcal{Y}} Y_i f(Y_i \mid M_i, X_i) dY_i \\
=&\int_{\mathcal{A}} \int_{\mathcal{Y}} Y_i f(Y_i \mid a, M_i, X_i) dY_i f(a|M_i, X_i) da \\
=&\int_{\mathcal{A}} \gamma_a(M_i, X_i) f(a|M_i, X_i) da < \infty.
\end{aligned}$$

Equality (a) follows from that

$$\begin{aligned}
&\sqrt{nh^{d_A}} \int_{\mathcal{M}\times\mathcal{X}\times\mathcal{Y}} f(a \mid Y_i, M_i, X_i) \left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i) Y_i f(Y_i, M_i, X_i) dY_i dM_i dX_i \\
=&\sqrt{nh^{d_A}} \int_{\mathcal{M}\times\mathcal{X}} \left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i) \left[\int_{\mathcal{Y}} Y_i f(Y_i \mid a, M_i, X_i) dY_i\right] f(a, M_i, X_i) dM_i dX_i \\
=&\sqrt{nh^{d_A}} \int_{\mathcal{M}\times\mathcal{X}} \left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i) \gamma_a(M_i, X_i) f(a, M_i, X_i) dM_i dX_i \\
=&\sqrt{nh^{d_A}} \int_{\mathcal{M}\times\mathcal{X}\times\mathcal{Y}} \left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i) \gamma_a(M_i, X_i) f(a, Y_i, M_i, X_i) dY_i dM_i dX_i \\
&\sqrt{nh^{d_A}} \int_{\mathcal{M}\times\mathcal{X}\times\mathcal{Y}} f(a \mid Y_i, M_i, X_i) \left[\hat{\lambda}_a(X_i) - \lambda_a(X_i)\right] R(M_i, X_i) \gamma_a(M_i, X_i) f(Y_i, M_i, X_i) dY_i dM_i dX_i
\end{aligned}$$

Proof for TR3

For Term (TR3), we have

$$\begin{aligned}
&\sqrt{nh^{d_A}} \mathbb{E}\big[\big(\hat{\eta}(X_i) - \eta(X_i)\big)\big(1 - K_h(A_i - a')\lambda_{a'}(X_i)\big)\big| O_\ell^c\big] \\
&= \sqrt{nh^{d_A}} \int_{X_i, A_i} \big(\hat{\eta}(X_i) - \eta(X_i)\big)\big(1 - K_h(A_i - a')\lambda_{a'}(X_i)\big) f(A_i, X_i) dA_i dX_i \\
&= \sqrt{nh^{d_A}} \int_{X_i} \big(\hat{\eta}(X_i) - \eta(X_i)\big)\big(1 - \big\{\int_{A_i} K_h(A_i - a') f(A_i \mid X_i) dA_i\big\}\lambda_{a'}(X_i)\big) f(X_i) dX_i \\
&\overset{(a)}{=} \sqrt{nh^{d_A}} \int_{X_i} \big(\hat{\eta}(X_i) - \eta(X_i)\big)\big(1 - f(a' \mid X_i)\lambda_{a'}(X_i)\big) f(X_i) dX_i \\
&\quad - \sqrt{nh^{d_A}} \int_{X_i} \big(\hat{\eta}(X_i) - \eta(X_i)\big) C h^2 f''(a' \mid X_i) \lambda_{a'}(X_i) f(X_i) dX_i \\
&\quad - \sqrt{nh^{d_A}} \int_{X_i} \big(\hat{\eta}(X_i) - \eta(X_i)\big) O(h^3) \lambda_{a'}(X_i) f(X_i) dX_i \\
&= o_p(1).
\end{aligned}$$

where (a) follows from Lemma 2, for the inequality we used boundedness of $f''(a|X)$ for all $a \in \mathcal{A}_0$, and the last equality follows form boundedness of the nuisance functions and that $nh^{d_A+4} \to C_h$, and Assumption 4(iv) (consistency of $\hat{\eta}(X_i)$).

Proof For TR4

The following Assumptions are utilized for the proof

- $nh^{d_A+4} \to C_h$
- $||\lambda_{a'}(X_i)f''(a' \mid M, X)||_\infty < \infty$
- Consistency of nuisance estimation

Now, demonstrating the bound for (TR4).

$$\mathbb{E}\big[\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\big\{K_h(A_i - a')\lambda_{a'}(X_i) - K_h(A_i - a)\lambda_a(X_i)R(M_i, X_i)\big\}\big]$$
$$= \mathbb{E}\big[\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\big\{K_h(A_i - a')\lambda_{a'}(X_i)\big\}\big] - \mathbb{E}\big[\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\big\{K_h(A_i - a)\lambda_a(X_i)R(M_i, X$$

Simplifying the above terms one at a time

$$\mathbb{E}\big[\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\big\{K_h(A_i - a')\lambda_{a'}(X_i)\big\}\big]$$
$$= \int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i)\left\{\int_{\mathcal{A}} K_h(A_i - a')f(A \mid M, X)da\right\} f(M, X)dmdx$$

From Lemma 2, $\int_{\mathcal{A}} K_h(A_i - a')f(A \mid M, X)da$ can be rewritten as $\int_{\mathcal{A}} = f(a' \mid M, X) + Ch^2 f''(a \mid M, X) + O_p(h^3)$

Plugging back into original integral gives:

$$\int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i)f(a' \mid M, X)f(M, X)dmdx$$
$$+ h^2 C \int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i)f''(a' \mid M, X)f(M, X)dmdx$$
$$+ O_p(h^2)\int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_{a'}(X_i)f(M, X)dmdx$$

Applying a similar approach to the second term, we get

$$-\int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_a(X_i)R(M_i, X_i)f(a \mid M, X)f(M, X)dmdx$$
$$-h^2 C \int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_a(X_i)R(M_i, X_i)f''(a \mid M, X)f(M, X)dmdx$$
$$-O_p(h^2)\int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)\lambda_a(X_i)R(M_i, X_i)f(M, X)dmdx$$

Claim:

$$\int_{\mathcal{X}}\int_{\mathcal{M}} \left(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\right)\left\{\lambda_{a'}(X_i) f(a' \mid M, X) - \lambda_a(X_i) R(M_i, X_i) f(a \mid M, X)\right\} f(M, X) dm dx = 0$$

Proof:

$$\lambda_{a'}(X_i) f(a' \mid M, X) = \frac{f(X)}{f(a', X)} \frac{f(a', M, X)}{f(M, X)}$$

Similarly

$$\begin{aligned}\lambda_a(X_i) R(M_i, X_i) f(a \mid M, X) &= \frac{f(X)}{f(a, X)} \frac{f(M, a', X)}{f(a', X)} \frac{f(a, X)}{f(M, a, X)} \frac{f(a, M, X)}{f(M, X)} \\ &= \frac{f(X)}{f(a', X)} \frac{f(M, a', X)}{f(M, X)}\end{aligned}$$

And so they cancel out. To bound the remaining terms, we follow a similar argument demonstrated below

$$\sqrt{nh^{d_A}} h^2 C \int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big) \lambda_{a'}(X_i) f''(a' \mid M, X) f(M, X) dm dx$$
$$\sqrt{nh^{d_A+4}} C \int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big) \lambda_{a'}(X_i) f''(a' \mid M, X) f(M, X) dm dx$$

By assumption, $nh^{d_A+4} \to C_h$. Applying Holder inequality to the integral with $p = 1$ and $q = \infty$

$$\leq \sqrt{nh^{d_A+4}} C \left\{\int_{\mathcal{X}}\int_{\mathcal{M}} |\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)| f(M, X) dm dx\right\} ||\lambda_{a'}(X_i) f''(a' \mid M, X)||_\infty$$

Since we assume the ∞-norm is bounded, the above becomes

$$= \sqrt{nh^{d_A+4}} C'' \left\{\int_{\mathcal{X}}\int_{\mathcal{M}} |\big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)| f(M, X) dm dx\right\}$$

Now, an application of Cauchy-Schwartz gives

$$\leq \sqrt{nh^{d_A+4}} C'' \left\{\int_{\mathcal{X}}\int_{\mathcal{M}} \big(\hat{\gamma}_a(M_i, X_i) - \gamma_a(M_i, X_i)\big)^2 f(M, X) dm dx\right\}^{\frac{1}{2}} \left\{\int_{\mathcal{X}}\int_{\mathcal{M}} f(M, X) dm dx\right\}^{\frac{1}{2}}$$

By the consistency assumption on our nuisance estimators and Slutskys theorem, the above term can be shown to be $o_p(1)$.

Proof For TR5

Finally, for term (TR5), we note that

$$
\begin{aligned}
&\sqrt{nh^{d_A}}\mathbb{E}\big[K_h(A_i-a')\big(\hat{\lambda}_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i,X_i)-\eta(X_i)\big\}\big|O_\ell^c\big] \\
&=\sqrt{nh^{d_A}}\int_{\mathcal{X},\mathcal{M},\mathcal{A}} K_h(A_i-a')\big(\hat{\lambda}_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i,X_i)-\eta(X_i)\big\}f(A_i,M_i,X_i)dA_idM_idX_i \\
&=\sqrt{nh^{d_A}}\int_{\mathcal{X},\mathcal{M}}\{\int_{\mathcal{A}} K_h(A_i-a')f(A_i\mid M_i,X_i)dA_i\} \\
&\quad\times\big(\hat{\lambda}_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i,X_i)-\eta(X_i)\big\}f(M_i,X_i)dM_idX_i \\
&\overset{(a)}{=}\sqrt{nh^{d_A}}\int_{\mathcal{X},\mathcal{M}}\big(\hat{\lambda}_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i,X_i)-\eta(X_i)\big\}f(a',M_i,X_i)dM_idX_i \\
&\quad+\sqrt{nh^{d_A}}\int_{\mathcal{X},\mathcal{M}}\big(\hat{\lambda}_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i,X_i)-\eta(X_i)\big\}Ch^2f''(a'\mid M_i,X_i)f(M_i,X_i)dM_idX_i \\
&\quad+\sqrt{nh^{d_A}}\int_{\mathcal{X},\mathcal{M}}\big(\hat{\lambda}_{a'}(X_i)-\lambda_{a'}(X_i)\big)\big\{\gamma_a(M_i,X_i)-\eta(X_i)\big\}O(h^3)f(M_i,X_i)dM_idX_i
\end{aligned}
$$

□

Next, for bias $B(a,a')$ and variance $V(a,a')$, given

$$\begin{aligned}
&\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a,a') - \psi_0(a,a')) \\
=&\sqrt{\frac{h^{d_A}}{n}}\sum_{l=1}^{L}\sum_{i\in I_\ell}\Big\{m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a'))\Big\} + o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
m(O_i;\alpha,\lambda,\gamma,\psi_0(a,a')) = &\frac{K_h(A_i-a)f(M_i \mid A=a',X_i)}{f(M_i \mid A=a,X_i)f(a \mid X_i)}\{Y_i - \mathbb{E}[Y \mid X_i, M_i, A=a]\} \\
&+\frac{K_h(A_i-a')}{f(a' \mid X_i)}\{\mathbb{E}[Y \mid X_i, M_i, A=a] - \eta(a,a',X_i)\} + \eta(a,a',X_i) - \psi_0(a,a')
\end{aligned}$$

we focus on

$$\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y - \mathbb{E}[Y \mid X,M,A=a]\} + \frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X,M,A=a] - \eta(a,a',X)\},$$

where $\eta(a,a',X) = \int \mathbb{E}[Y \mid X, M=m, A=a]f(m \mid A=a',X)dm = \mathbb{E}\{\mathbb{E}[Y \mid X, A=a]|A=a',X\}$.

Expectation Part 1

$$\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y - \mathbb{E}[Y \mid X,M,A=a]\}$$

From $\mathbb{E}\{\mathbb{E}[\gamma(X,M,A)|X,M]\} = \mathbb{E}\{\mathbb{E}[Y|X,M]\}$ and $\gamma(X,M,A) = \mathbb{E}(Y|X,M,A)$, expectation of the first term

$$\begin{aligned}
&\mathbb{E}\left[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y - \mathbb{E}[Y \mid X,M,A=a]\}\right] \\
=&\mathbb{E}\left\{\mathbb{E}\left[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y - \mathbb{E}[Y \mid X,M,A=a]\}\middle| X,M\right]\right\} \\
=&\mathbb{E}\left\{\frac{f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\mathbb{E}\left[K_h(A-a)(\gamma(X,M,A) - \gamma(X,M,a))\middle| X,M\right]\right\}.
\end{aligned}$$

The inner product further expands as follows,

$$
\begin{aligned}
&\mathbb{E}\Big[K_h(A-a)(\gamma(X,M,A)-\gamma(X,M,a))\Big|X,M\Big]\\
=&\int_{\mathcal{A}} K_h(A-a)(\gamma(X,M,A)-\gamma(X,M,a))f(A|X,M)dA\\
=&\int_{\mathcal{A}}\Big[\prod_{j=1}^{d_A}\frac{1}{h}k\Big(\frac{A_j-a}{h}\Big)\Big](\gamma(X,M,A)-\gamma(X,M,a))f(A|X,M)dA\\
=&\int k(u)(\gamma(X,M,a+uh)-\gamma(X,M,a))f(a+uh|X,M)du\\
=&\int k(u_1)\cdots k(u_{d_A})\Big(\sum_{j=1}^{d_A}u_jh\partial_{a_j}\gamma(X,M,a)+\frac{u_j^2h^2}{2}\partial_{a_j}^2\gamma(X,M,a)\Big)\\
&\times\Big(f(a|X,M)+\sum_{j=1}^{d_A}u_jh\partial_{a_j}f(a|X,M)+\frac{u_j^2h^2}{2}\partial_{a_j}^2 f(a|X,M)\Big)du_1\cdots du_{d_A}+O(h^3)\\
=&h^2\int u^2k(u)du\Big(\sum_{j=1}^{d_A}\partial_{a_j}\gamma(X,M,a)\partial_{a_j}f(a|X,M)+\frac{1}{2}[\partial_{a_j}^2\gamma(X,M,a)]f(a|X,M)\Big)+O(h^3)
\end{aligned}
$$

for all X,M in respective range.

$$
\begin{aligned}
&\mathbb{E}\Big\{\frac{f(M\mid A=a',X)}{f(M\mid A=a,X)f(a\mid X)}\mathbb{E}\Big[K_h(A-a)(\gamma(X,M,A)-\gamma(X,M,a))\Big|X,M\Big]\Big\}\\
=&h^2\int u^2k(u)du\\
&\times\mathbb{E}\Big[\frac{f(M\mid A=a',X)}{f(M\mid A=a,X)}\Big(\partial_a\gamma(X,M,a)\frac{\partial_a f(a|X,M)}{f(a\mid X)}+\frac{1}{2}\frac{f(a\mid X,M)}{f(a\mid X)}\partial_a^2\gamma(X,M,a)\Big)\Big]+O(h^3)
\end{aligned}
$$

Expectation Part 2

$$
\frac{K_h(A-a')}{f(a'\mid X)}\{\mathbb{E}[Y\mid X,M,A=a]-\eta(a,a',X)\}
$$

$$
\begin{aligned}
&\mathbb{E}\bigg[\frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)\}\bigg]\\
=&\mathbb{E}\bigg\{\mathbb{E}\bigg[\frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)\}\bigg| X, M\bigg]\bigg\}\\
=&\mathbb{E}\bigg\{\frac{1}{f(a' \mid X)}\mathbb{E}\bigg[K_h(A-a')\{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)\}\bigg| X, M\bigg]\bigg\}\\
=&\mathbb{E}\bigg\{\frac{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)}{f(a' \mid X)}\mathbb{E}[K_h(A-a')\bigg| X, M]\bigg\}
\end{aligned}
$$

The inner expectation becomes

$$
\begin{aligned}
&\mathbb{E}\bigg[K_h(A-a')\bigg| X, M\bigg]\\
=&\int_{\mathcal{A}}\bigg[\prod_{j=1}^{d_A}\frac{1}{h}k\Big(\frac{A_j-a'}{h}\Big)\bigg]f(A|X,M)dA\\
=&\int k(u_1)\cdots k(u_{d_A})\bigg(f(a'|X,M)+\sum_{j=1}^{d_A}u_j h\partial_{a_j}f(a'|X,M)+\frac{u_j^2h^2}{2}\partial_{a_j}^2 f(a'|X,M)\bigg)du_1\cdots du_{d_A}+O(h^3)\\
=&f(a'|X,M)+\frac{1}{2}h^2\int u^2k(u)du\sum_{j=1}^{d_A}\partial_{a_j}^2 f(a'|X,M)+O(h^3)
\end{aligned}
$$

The expectation of part 2

$$
\begin{aligned}
&\mathbb{E}\bigg\{\frac{1}{f(a'|X)}\{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)\}\\
&\times\bigg(f(a'|X,M)+\frac{1}{2}h^2\int u^2k(u)du\partial_a^2 f(a'|X,M)\bigg)\bigg\}+O(h^3)\\
=&\mathbb{E}\bigg\{\{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)\}\\
&\times\bigg(\frac{f(a'|X,M)}{f(a'|X)}+\frac{1}{2}h^2\int u^2k(u)du\frac{\partial_a^2 f(a'|X,M)}{f(a'|X)}\bigg)\bigg\}+O(h^3)\\
=&h^2\int u^2k(u)du\mathbb{E}\bigg\{\{\mathbb{E}[Y \mid X, M, A=a]-\eta(a,a',X)\}\frac{1}{2}\frac{\partial_a^2 f(a'|X,M)}{f(a'|X)}\bigg\}+O(h^3)
\end{aligned}
$$

from having the first term in this expectation equal to zero,

$$\begin{aligned}
&\mathbb{E}\left\{\{\mathbb{E}[Y \mid X, M, A=a] - \eta(a,a',X)\}\frac{f(a'|X,M)}{f(a'|X)}\right\} \\
&= \int\int \{\mathbb{E}[Y \mid X, M, A=a] - \eta(a,a',X)\}\frac{f(a'|X,M)}{f(a'|X)} f(M,X) dM dX \\
&= \int\int \{\mathbb{E}[Y \mid X, M, A=a] - \eta(a,a',X)\}\frac{f(A=a',X,M)}{f(A=a',X)} f(X) dM dX \\
&= \int\int \{\mathbb{E}[Y \mid X, M, A=a] - \eta(a,a',X)\} f(M|A=a',X) dM f(X) dX \\
&= \int \{\eta(a,a',X) - \eta(a,a',X)\} f(X) dX = 0
\end{aligned}$$

Hence,

$$\begin{aligned}
&B(a,a') = \\
&\quad h^2 \int u^2 k(u) du \\
&\quad \times \mathbb{E}\left\{\frac{f(M \mid A=a',X)}{f(M \mid A=a,X)}\left(\partial_a \gamma(X,M,a)\frac{\partial_a f(a|X,M)}{f(a \mid X)} + \frac{1}{2}\frac{f(a \mid X,M)}{f(a \mid X)}\partial_a^2 \gamma(X,M,a)\right)\right. \\
&\quad \left. + \{\mathbb{E}[Y \mid X, M, A=a] - \eta(a,a',X)\}\frac{1}{2}\frac{\partial_a^2 f(a'|X,M)}{f(a'|X)}\right\} + O(h^3)
\end{aligned}$$

Variance

$$\begin{aligned}
h^{d^A} \times var&\left\{\frac{K_h(A-a) f(M \mid A=a',X)}{f(M \mid A=a,X) f(a \mid X)}\{Y - \mathbb{E}[Y \mid X, M, A=a]\}\right. \\
&+ \frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X, M, A=a] - \eta(a,a',X)\} \\
&\left. + \eta(a,a',X) - \theta_0\right\}
\end{aligned}$$

To start with, we look at the second order expectation

$$
\begin{aligned}
&\mathbb{E}\Bigg\{\Bigg[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y-\mathbb{E}[Y \mid X,M,A=a]\} \\
&+\frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X,M,A=a]-\eta(a,a',X)\}\Bigg]^2\Bigg\} \\
=&\mathbb{E}\Bigg\{\Bigg[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y-\mathbb{E}[Y \mid X,M,A=a]\}\Bigg]^2\Bigg\} \\
&+\mathbb{E}\Bigg\{\Bigg[\frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X,M,A=a]-\eta(a,a',X)\}\Bigg]^2\Bigg\} \\
&+2\mathbb{E}\Bigg\{\Bigg[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y-\mathbb{E}[Y \mid X,M,A=a]\}\Bigg] \\
&\quad\times\Bigg[\frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X,M,A=a]-\eta(a,a',X)\}\Bigg]\Bigg\}
\end{aligned}
$$

Variance Part 1

$$
\begin{aligned}
&\mathbb{E}\left\{\left[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y-\mathbb{E}[Y \mid X,M,A=a]\}\right]^2\right\}\\
=&\mathbb{E}\left\{\mathbb{E}\left\{\left[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y-\mathbb{E}[Y \mid X,M,A=a]\}\right]^2\middle|X,M\right\}\right\}\\
=&\mathbb{E}\left\{\frac{f(M \mid A=a',X)^2}{f(M \mid A=a,X)^2f(a \mid X)^2}\mathbb{E}\left\{K_h(A-a)^2(Y-\mathbb{E}[Y \mid X,M,A=a])^2\middle|X,M\right\}\right\}\\
=&\mathbb{E}\left\{\frac{f(M \mid A=a',X)^2}{f(M \mid A=a,X)^2f(a \mid X)^2}\right.\\
&\left.\times\mathbb{E}\left\{K_h(A-a)^2\mathbb{E}\{(Y-\mathbb{E}[Y \mid X,M,A=a])^2|X,M,A\}\middle|X,M\right\}\right\}\\
=&\mathbb{E}\left\{\frac{f(M \mid A=a',X)^2}{f(M \mid A=a,X)^2f(a \mid X)^2}\right.\\
&\left.\times\mathbb{E}\left\{K_h(A-a)^2\left[var(Y|X,M,A)+\gamma(X,M,A)^2-2\gamma(X,M,A)\gamma(X,M,a)+\gamma(X,M,a)^2\right]\middle|X,M\right\}\right\}\\
=&\mathbb{E}\left\{\frac{f(M \mid A=a',X)^2}{f(M \mid A=a,X)^2f(a \mid X)^2}\right.\\
&\left.\times\mathbb{E}\left\{K_h(A-a)^2\left[var(Y|X,M,A)+(\gamma(X,M,A)-\gamma(X,M,a))^2\right]\middle|X,M\right\}\right\}
\end{aligned}
$$

The inner expectation

$$
\begin{aligned}
&\mathbb{E}\bigg\{K_h(A-a)^2\Big[var(Y|X,M,A)+(\gamma(X,M,A)-\gamma(X,M,a))^2\Big]\bigg|X,M\bigg\}\\
=&\int\bigg[\prod_{j=1}^{d_A}\frac{1}{h^2}k\Big(\frac{A_j-a}{h}\Big)^2\bigg]\Big[var(Y|X,M,A)+(\gamma(X,M,A)-\gamma(X,M,a))^2\Big]f(A|X,M)dA\\
=&\frac{1}{h^{d_A}}\int k(u)^2\\
&\times\Big[var(Y|X,M,a+uh)+(\gamma(X,M,a+uh)-\gamma(X,M,a))^2\Big]f(a+uh|X,M)du\\
=&\frac{1}{h^{d_A}}\int k(u_1)^2\cdots k(u_{d_A})^2\\
&\times\bigg[var(Y|X,M,a)+\sum_{j=1}^{d_A}u_jh\partial_{a_j}var(Y|X,M,a)+\frac{u_j^2h^2}{2}\partial_{a_j}^2var(Y|X,M,a)+O(h^3)\\
&\qquad+\bigg(\sum_{j=1}^{d_A}u_jh\partial_{a_j}\gamma(X,M,a)+\frac{u_j^2h^2}{2}\partial_{a_j}^2\gamma(X,M,a)+O(h^3)\bigg)^2\bigg]\\
&\times\bigg(f(a|X,M)+\sum_{j=1}^{d_A}u_jh\partial_{a_j}f(a|X,M)+\frac{u_j^2h^2}{2}\partial_{a_j}^2f(a|X,M)+O(h^3)\bigg)du_1\cdots du_{d_A}\\
=&\frac{1}{h^{d_A}}\bigg\{\int k(u)^2du\times var(Y|X,M,a)f(a|X,M)\\
&+h^2\int u^2k(u)^2du\\
&\qquad\times\bigg\{\Big[\frac{1}{2}\partial_a^2var(Y|X,M,a)+(\partial_a\gamma(X,M,a))^2\Big]f(a|X,M)+\partial_avar(Y|X,M,a)\partial_af(a|X,M)\\
&\qquad+\frac{1}{2}var(Y|X,M,a)\partial_a^2f(a|X,M)\bigg\}+O(h^4)\bigg\}
\end{aligned}
$$

Hence, the part 1 of variance

$$
\begin{aligned}
&\mathbb{E}\left\{\left[\frac{K_h(A-a)f(M\mid a',X)}{f(M\mid a,X)f(a\mid X)}\{Y-\mathbb{E}[Y\mid X,M,a]\}\right]^2\right\}\\
=&\mathbb{E}\left\{\frac{f(M\mid a',X)^2}{f(M\mid a,X)^2f(a\mid X)^2}\right.\\
&\times\left[\frac{1}{h^{d_A}}\int k(u)^2du\times var(Y|X,M,a)f(a|X,M)\right.\\
&\qquad+h^{2-d_A}\int u^2k(u)^2du\times\left(\left[\frac{1}{2}\partial_a^2 var(Y|X,M,a)+(\partial_a\gamma(X,M,a))^2\right]f(a|X,M)\right.\\
&\qquad\left.\left.\left.+\partial_a var(Y|X,M,a)\partial_a f(a|X,M)+\frac{1}{2}var(Y|X,M,a)\partial_a^2 f(a|X,M)\right)\right]\right\}+O(h^{4-d_A})\\
=&\frac{1}{h^{d_A}}\int k(u)^2du\times\mathbb{E}\left\{\frac{f(M\mid a',X)^2f(a|X,M)}{f(M\mid a,X)^2f(a\mid X)^2}var(Y|X,M,a)\right\}\\
&+h^{2-d_A}\int k(u)^2u^2du\times\mathbb{E}\left\{\frac{f(M\mid a',X)^2f(a|X,M)}{f(M\mid a,X)^2f(a\mid X)^2}\left[\frac{1}{2}\partial_a^2 var(Y|X,M,a)+(\partial_a\gamma(X,M,a))^2\right]\right.\\
&\qquad\left.+\frac{f(M\mid a',X)^2}{f(M\mid a,X)^2f(a\mid X)^2}\left[\partial_a var(Y|X,M,a)\partial_a f(a|X,M)+\frac{1}{2}var(Y|X,M,a)\partial_a^2 f(a|X,M)\right]\right\}\\
&+O(h^{4-d_A})
\end{aligned}
$$

Variance Part 2

$$
\begin{aligned}
&\mathbb{E}\left\{\left[\frac{K_h(A-a')}{f(a'\mid X)}\Big(\mathbb{E}(Y\mid X,M,A=a)-\eta(a,a',X)\Big)\right]^2\right\}\\
=&\mathbb{E}\left\{\frac{1}{f(a'\mid X)^2}\mathbb{E}\left[K_h(A-a')^2\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2\Big|X\right]\right\}
\end{aligned}
$$

The inner expectation

$$
\begin{aligned}
&\mathbb{E}\Big[K_h(A-a')^2\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2\Big|X\Big]\\
=&\int\int\Big[\prod_{j=1}^{d_A}\frac{1}{h^2}k\Big(\frac{A_j-a'}{h}\Big)^2\Big]\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2 f(A|X)f(M|A,X)dAdM\\
=&\frac{1}{h^{d_A}}\int\int k(u)^2\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2 f(a'+uh|X)f(M|a'+uh,X)dudM\\
=&\frac{1}{h^{d_A}}\int\int k(u_1)^2\cdots k(u_{d_A})^2\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2\\
&\times\Big(f(a'|X)+\sum_{j=1}^{d_A}u_jh\partial_{a_j}f(a'|X)+\frac{u_j^2h^2}{2}\partial_{a_j}^2 f(a'|X)+O(h^3)\Big)\\
&\times\Big(f(M|a',X)+\sum_{j=1}^{d_A}u_jh\partial_{a_j}f(M|a',X)+\frac{u_j^2h^2}{2}\partial_{a_j}^2 f(M|a',X)+O(h^3)\Big)du_1\cdots du_{d_A}dM\\
=&\frac{1}{h^{d_A}}\int k(u)^2du\times var[E(Y|X,M,a)|X,a']f(a'|X)\\
&+h^{2-d_A}\int k(u)^2u^2du\times\Big\{\frac{1}{2}var[E(Y|X,M,a)|X,a']\partial_a^2 f(a'|X)\\
&+\int\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2\times\Big[\frac{1}{2}f(a'|X)\partial_a^2 f(M|X,a')+\partial_a f(a'|X)\partial_a f(M|X,a')\Big]dM\Big\}\\
&+O(h^{4-d_A})
\end{aligned}
$$

the last equation is from

$$
\begin{aligned}
var[E(Y|X,M,a)|X,a'] =&\mathbb{E}\Big\{\Big[E(Y|X,M,a)-\eta(a,a',X)\Big]^2|X,a'\Big\}\\
=&\int\Big[E(Y|X,M,a)-\eta(a,a',X)\Big]^2 f(M|X,a')dM.
\end{aligned}
$$

Hence, the part 2 of variance

$$\mathbb{E}\left\{\frac{1}{f(a' \mid X)^2}\mathbb{E}\left[K_h(A-a')^2\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2\Big|X\right]\right\}$$

$$=\frac{1}{h^{d_A}}\int k(u)^2du \times \mathbb{E}\left\{\frac{1}{f(a'|X)}var[E(Y|X,M,a)|X,a']\right\}$$

$$+ h^{2-d_A}\int k(u)^2u^2du \times \frac{1}{2}\mathbb{E}\left\{\frac{\partial_a^2 f(a'|X)}{f(a'|X)^2}var[E(Y|X,M,a)|X,a']\right\}$$

$$+ h^{2-d_A}\int k(u)^2u^2du \times$$

$$\mathbb{E}\left\{\Big(\gamma(X,M,a)-\eta(a,a',X)\Big)^2 \times \Big[\frac{\partial_a^2 f(M|X,a')}{2f(a'|X)f(M|X,A)} + \frac{\partial_a f(a'|X)\partial_a f(M|X,a')}{f(a'|X)^2 f(M|X,A)}\Big]\right\}$$

$$+ O(h^{4-d_A})$$

Variance Part 3

$$\mathbb{E}\left\{\left[\frac{K_h(A-a)f(M \mid A=a',X)}{f(M \mid A=a,X)f(a \mid X)}\{Y-\mathbb{E}[Y \mid X,M,A=a]\}\right]\right.$$

$$\left.\times\left[\frac{K_h(A-a')}{f(a' \mid X)}\{\mathbb{E}[Y \mid X,M,A=a]-\eta(a,a',X)\}\right]\right\}$$

$$=\mathbb{E}\left\{\frac{K_h(A-a)K_h(A-a')}{f(a|X)f(a'|X)}\frac{f(M|a',X)}{f(M|a,X)}\Big[Y-\gamma(X,M,a)\Big]\Big[\gamma(X,M,a)-\eta(a,a',X)\Big]\right\}$$

$$=\mathbb{E}\left\{\frac{1}{f(a|X)f(a'|X)}\frac{f(M|a',X)}{f(M|a,X)}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big]\right.$$

$$\left.\times\,\mathbb{E}\Big\{K_h(A-a)K_h(A-a')\Big[Y-\gamma(X,M,a)\Big]\Big|X,M\Big\}\right\}$$

$$=\mathbb{E}\left\{\frac{1}{f(a|X)f(a'|X)}\frac{f(M|a',X)}{f(M|a,X)}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big]\right.$$

$$\left.\times\,\mathbb{E}\Big\{K_h(A-a)K_h(A-a')\Big[\gamma(X,M,A)-\gamma(X,M,a)\Big]\Big|X,M\Big\}\right\}$$

The inner expectation

$$
\begin{aligned}
&\mathbb{E}\Big\{K_h(A-a)K_h(A-a')\Big[\gamma(X,M,A)-\gamma(X,M,a)\Big]\Big|X,M\Big\}\\
=&\int\Big[\prod_{j=1}^{d_A}\frac{1}{h^2}k\Big(\frac{A_j-a}{h}\Big)k\Big(\frac{A_j-a'}{h}\Big)\Big]\Big[\gamma(X,M,A)-\gamma(X,M,a)\Big]f(A|X,M)dA\\
=&\int\frac{1}{h^{d_A}}k(u_1)\cdots k(u_{d_A})k(u_1+\frac{a-a'}{h})\cdots k(u_{d_A}+\frac{a-a'}{h})\\
&\times\Big[\sum_{j=1}^{d_A}u_jh\partial_{a_j}\gamma(X,M,a)+\frac{u_j^2h^2}{2}\partial^2_{a_j}\gamma(X,M,a)+\frac{u_j^3h^3}{6}\partial^3_{a_j}\gamma(X,M,a)+O(h^4)\Big]\\
&\times\Big[f(a|X,M)+\sum_{j=1}^{d_A}u_jh\partial_{a_j}f(a|X,M)+\frac{u_j^2h^2}{2}\partial^2_{a_j}f(a|X,M)+O(h^3)\Big]du_1\cdots du_{d_A}\\
=&h^{1-d_A}\int k(u)k(u+\frac{a-a'}{h})udu\times f(a|X,M)\partial_a\gamma(X,M,a)\\
&+h^{2-d_A}\int k(u)k(u+\frac{a-a'}{h})u^2du\times\Big[\frac{1}{2}f(a|X,M)\partial_a^2\gamma(X,M,a)+\partial_af(a|X,M)\partial_a\gamma(X,M,a)\Big]\\
&+h^{3-d_A}\int k(u)k(u+\frac{a-a'}{h})u^3du\times\\
&\qquad\Big[\frac{1}{6}f(a|X,M)\partial_a^3\gamma(X,M,a)+\frac{1}{2}\partial_af(a|X,M)\partial_a^2\gamma(X,M,a)+\frac{1}{2}\partial_a^2f(a|X,M)\partial_a\gamma(X,M,a)\Big]\\
&+O(h^{4-d_A})
\end{aligned}
$$

Hence, the part 3 of variance

$$
\begin{aligned}
&\mathbb{E}\bigg\{\frac{1}{f(a|X)f(a'|X)}\frac{f(M|a',X)}{f(M|a,X)}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big] \\
&\qquad \times \mathbb{E}\Big\{K_h(A-a)K_h(A-a')\Big[\gamma(X,M,A)-\gamma(X,M,a)\Big]\Big|X,M\Big\}\bigg\} \\
=&\mathbb{E}\bigg\{\frac{1}{f(a|X)f(a'|X)}\frac{f(M|a',X)}{f(M|a,X)}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big]\times \\
&\qquad \Big[h^{1-d_A}\int k(u)k(u+\frac{a-a'}{h})udu \times f(a|X,M)\partial_a\gamma(X,M,a) \\
&\qquad + h^{2-d_A}\int k(u)k(u+\frac{a-a'}{h})u^2du \times \Big[\frac{1}{2}f(a|X,M)\partial_a^2\gamma(X,M,a)+\partial_a f(a|X,M)\partial_a\gamma(X,M,a)\Big] \\
&\qquad + h^{3-d_A}\int k(u)k(u+\frac{a-a'}{h})u^3du\times \\
&\qquad \Big[\frac{1}{6}f(a|X,M)\partial_a^3\gamma(X,M,a)+\frac{1}{2}\partial_a f(a|X,M)\partial_a^2\gamma(X,M,a)+\frac{1}{2}\partial_a^2 f(a|X,M)\partial_a\gamma(X,M,a)\Big]\Big]\bigg\} \\
&+O(h^{4-d_A}) \\
=&h^{1-d_A}\int k(u)k(u+\frac{a-a'}{h})udu \times \mathbb{E}\Big\{\frac{f(a'|M,X)}{f(a'|X)^2}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big]\partial_a\gamma(X,M,a)\Big\} \\
&+h^{2-d_A}\int k(u)k(u+\frac{a-a'}{h})u^2du \\
&\qquad \times \mathbb{E}\Big\{\frac{f(a'|M,X)}{f(a'|X)^2}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big]\Big[\frac{1}{2}\partial_a^2\gamma(X,M,a)+\frac{\partial_a f(a|X,M)\partial_a\gamma(X,M,a)}{f(a|M,X)}\Big]\Big\} \\
&+h^{3-d_A}\int k(u)k(u+\frac{a-a'}{h})u^3du \times \mathbb{E}\Big\{\frac{f(a'|M,X)}{f(a'|X)^2}\Big[\gamma(X,M,a)-\eta(a,a',X)\Big] \\
&\qquad \times \Big[\frac{1}{6}\partial_a^3\gamma(X,M,a)+\frac{\partial_a f(a|X,M)\partial_a^2\gamma(X,M,a)}{2f(a|M,X)}+\frac{\partial_a^2 f(a|X,M)\partial_a\gamma(X,M,a)}{2f(a|M,X)}\Big]\Big\} \\
&+O(h^{4-d_A})
\end{aligned}
$$

Summarizing the three parts, the variance can be written as

$$
\begin{aligned}
V(a,a') =& h^{d_A} \times \Bigg\{ \frac{1}{h^{d_A}} \int k(u)^2 du \times E_1 + h^{2-d_A} \int k(u)^2 u^2 du \times E_2 \\
& + h^{1-d_A} \int k(u)k(u + \frac{a-a'}{h}) u du \times E_3 + h^{2-d_A} \int k(u)k(u + \frac{a-a'}{h}) u^2 du \times E_4 \\
& + h^{3-d_A} \int k(u)k(u + \frac{a-a'}{h}) u^3 du \times E_5 + O(h^{4-d_A}) \Bigg\} \\
=& \int k(u)^2 du \times E_1 + h^2 \int k(u)^2 u^2 du \times E_2 \\
& + h \int k(u)k(u + \frac{a-a'}{h}) u du \times E_3 + h^2 \int k(u)k(u + \frac{a-a'}{h}) u^2 du \times E_4 \\
& + h^3 \int k(u)k(u + \frac{a-a'}{h}) u^3 du \times E_5 + O(h^4)
\end{aligned}
$$

where

$$
\begin{aligned}
E_1 =& \mathbb{E}\Bigg\{ \frac{f(M \mid a', X)^2 f(a|X,M)}{f(M \mid a, X)^2 f(a \mid X)^2} var(Y|X,M,a) + \frac{1}{f(a'|X)} var[E(Y|X,M,a)|X,a'] \Bigg\} \\
E_2 =& \mathbb{E}\Bigg\{ \frac{f(M \mid a', X)^2 f(a|X,M)}{f(M \mid a, X)^2 f(a \mid X)^2} \Big[\frac{1}{2}\partial_a^2 var(Y|X,M,a) + (\partial_a \gamma(X,M,a))^2 \Big] \\
& + \frac{f(M \mid a', X)^2}{f(M \mid a, X)^2 f(a \mid X)^2} \Big[\partial_a var(Y|X,M,a) \partial_a f(a|X,M) + \frac{1}{2} var(Y|X,M,a) \partial_a^2 f(a|X,M) \Big] \\
& + \frac{\partial_a^2 f(a'|X)}{2 f(a'|X)^2} var[E(Y|X,M,a)|X,a'] \\
& + \Big(\gamma(X,M,a) - \eta(a,a',X) \Big)^2 \times \Big[\frac{\partial_a^2 f(M|X,a')}{2 f(a'|X) f(M|X,A)} + \frac{\partial_a f(a'|X) \partial_a f(M|X,a')}{f(a'|X)^2 f(M|X,A)} \Big] \Bigg\} \\
E_3 =& \mathbb{E}\Bigg\{ \frac{f(a'|M,X)}{f(a'|X)^2} \Big[\gamma(X,M,a) - \eta(a,a',X) \Big] \partial_a \gamma(X,M,a) \Bigg\} \\
E_4 =& \mathbb{E}\Bigg\{ \frac{f(a'|M,X)}{f(a'|X)^2} \Big[\gamma(X,M,a) - \eta(a,a',X) \Big] \Big[\frac{1}{2}\partial_a^2 \gamma(X,M,a) + \frac{\partial_a f(a|X,M) \partial_a \gamma(X,M,a)}{f(a|M,X)} \Big] \Bigg\} \\
E_5 =& \mathbb{E}\Bigg\{ \frac{f(a'|M,X)}{f(a'|X)^2} \Big[\gamma(X,M,a) - \eta(a,a',X) \Big] \\
& \times \Big[\frac{1}{6}\partial_a^3 \gamma(X,M,a) + \frac{\partial_a f(a|X,M) \partial_a^2 \gamma(X,M,a)}{2 f(a|M,X)} + \frac{\partial_a^2 f(a|X,M) \partial_a \gamma(X,M,a)}{2 f(a|M,X)} \Big] \Bigg\}
\end{aligned}
$$

Hence, as $h \to 0$, the asymptotic variance $V(a,a')$ converges to $\int k(u)^2 du \times E_1$.

Normality

Write $IF_k = m(O_k; \alpha, \lambda, \gamma, \psi_0(a, a'))$, raw moments as $\mu_\ell = \mathbb{E}(IF_k^l)$, and central moments as $\mu'_\ell = \mathbb{E}\Big\{(IF_k - \mathbb{E}(IF_k))^l\Big\}$. Note that IF_k is also a function of h. For normality, we want to show that IF_k satisfies the Lyapunov's condition for some $\delta > 0$,

$$\Big\{\sum_{k=1}^{n} var(IF_k)\Big\}^{-(2+\delta)/2} \sum_{k=1}^{n} \mathbb{E}\Big[\Big|IF_k - \mathbb{E}(IF_k)\Big|^{2+\delta}\Big] \to 0$$

Previous derivation shows that $\mathbb{E}(IF_k) = \mu_1 = O(h^2)$ and $var(IF_k) = \mu'_2 = O(\frac{1}{h^{d_A}})$. From standard algebra, we can show that $\mu'_\ell = \mathbb{E}\Big\{(IF_k - \mathbb{E}(IF_k))^l\Big\} = O(\frac{1}{h^{(l-1)d_A}})$ and taking δ to be an even number,

$$\mathbb{E}\Big[\Big|IF_k - \mathbb{E}(IF_k)\Big|^{2+\delta}\Big] = \mathbb{E}\Big[\Big(IF_k - \mathbb{E}(IF_k)\Big)^{2+\delta}\Big] = O\Big(\frac{1}{h^{(1+\delta)d_A}}\Big).$$

Hence,

$$\begin{aligned}&\Big\{\sum_{k=1}^{n} var(IF_k)\Big\}^{-(2+\delta)/2} \sum_{k=1}^{n} \mathbb{E}\Big[\Big|IF_k - \mathbb{E}(IF_k)\Big|^{2+\delta}\Big] \\ =&\Big\{nO\Big(\frac{1}{h^{d_A}}\Big)\Big\}^{-(2+\delta)/2} nO\Big(\frac{1}{h^{(1+\delta)d_A}}\Big) \\ =&O\Big(\frac{1}{\sqrt{(nh^{d_A})^{\delta}}}\Big) = o(1).\end{aligned}$$

Thus, the Lyapunov's condition is satisfied and by CLT,

$$\frac{\sum_{k=1}^{n} IF_k - n * \mathbb{E}(IF_k)}{\sqrt{n * var(IF_k)}} \sim N(0, 1) \qquad \text{as } n \to \infty.$$

Therefore, $\sqrt{nh^{d_A}}(\hat{\psi}^{TR}(a, a') - \psi_0(a, a') - B(a, a')) \sim N(0, V(a, a'))$ as $n \to \infty$.

6.2 Proof for Uniform Convergence

R1-1/R1-2 type terms: 2,3,7,9,10,12,13,14; R2 type terms: 1,4,5,6,8,11

- kernel $k(\cdot)$ is bounded
- $\sup_{a',a\in\mathcal{A}_0} \sqrt{nh^{d_A}}(\int_{\mathcal{X}\times\mathcal{M}}(\hat{\lambda}(a',m,x)-\lambda(a',m,x))^2 f(m,a',x)dmdx)^{1/2}(\int_{\mathcal{X}\times\mathcal{M}}(\hat{\gamma}(a,m,x)-\gamma(a,m,x))^2 f(m,a,x)dmdx)^{1/2}\xrightarrow{p}0$
- There exist $A_{1n}\to 0$,$A_{2n}\to 0$, and $A_{3n}\to 0$ such that

  $$\sup_{a\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}} |\hat{\gamma}_l(a,M_i,X_i)-\gamma(a,M_i,X_i)| = O_p(A_{1n}),$$

  $$\sup_{a\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}} |\hat{\lambda}_l(a,M_i,X_i)-\lambda(a,M_i,X_i)| = O_p(A_{2n})$$

  and

  $$\sup_{a,a'\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}} |\hat{R}_l(a',a,M_i,X_i)-R(a',a,M_i,X_i)| = O_p(A_{3n})$$

  where $R_\ell(a',a,m,x)=\alpha(a',m,x)/\alpha(a,m,x)$.
- $\hat{\gamma}_l(a,m,x)$ and $\hat{\lambda}_l(a,m,x)$ are Lipschitz continuous in $a\in\mathcal{A}_0$, for any $m\in\mathcal{M}$ and $x\in\mathcal{X}$.

Focus on term (CS5) in Theorem 1's proof,

$$\sqrt{\frac{h^{d_A}}{n}}\sum_{i\in\ell}\left\{K_h(A_i-a')\left[\hat{\lambda}(a',M_i,X_i)-\lambda(a',M_i,X_i)\right]\left[\hat{\gamma}(a,M_i,X_i)-\gamma(a,M_i,X_i)\right]\right\} \quad (8)$$

and show that it is $o_p(1)$ uniformly over $(a,a')\in\mathcal{A}_0$. Write $\hat{\lambda}^l_a(X_i)$ and $\hat{\gamma}^l_a(M_i,X_i)$ as the nuisance estimators that use O^c_ℓ for estimation and applied on subject i where $i\in O_\ell$. Corresponding to the term under consideration, we define

$$\Delta_{i\ell}(a',a)=\left[\hat{\lambda}_\ell(a',M_i,X_i)-\lambda(a',M_i,X_i)\right]\left[\hat{\gamma}_\ell(a,M_i,X_i)-\gamma(a,M_i,X_i)\right],$$
$$\hat{g}(a',a)=n^{-1}\sum_{l=1}^{L}\sum_{i\in I_\ell}K_h(A_i-a')\Delta_{i\ell}(a',a),\ \text{and}$$
$$W_{i\ell}(a',a)=K_h(A_i-a')\Delta_{i\ell}(a',a)-\mathbb{E}\left[K_h(A_i-a')\Delta_{i\ell}(a',a)\right].$$

Lemma 3 (Supremum of g) $\sup_{a',a\in\mathcal{A}_0}\mathbb{E}[\hat{g}(a',a)]=o_p(\sqrt{1/(nh^{d_A})})$

Proof:

$$\sqrt{nh^{d_A}}\mathbb{E}[\hat{g}(a',a)] = \sqrt{nh^{d_A}}\mathbb{E}[K_h(A_i - a')\Delta_{i\ell}(a',a)]$$

$$= \sqrt{nh^{d_A}} \int_{\mathcal{X}\times\mathcal{M}\times\mathcal{A}} K_h(A_i - a')(\hat{\lambda}_\ell(a', M_i, X_i) - \lambda(a', M_i, X_i))$$
$$(\hat{\gamma}_\ell(a, M_i, X_i) - \gamma(a, M_i, X_i))f(A_i, M_i, X_i)dX_i dM_i dA_i$$

$$= \sqrt{nh^{d_A}} \int_{\mathcal{X}\times\mathcal{M}} [f(a'|M_i, X_i) + O_p(h^2)](\hat{\lambda}(a', M_i, X_i) - \lambda(a', M_i, X_i))$$
$$(\hat{\gamma}(a, M_i, X_i) - \gamma(a, M_i, X_i))f(M_i, X_i)dX_i dM_i$$

$$\leq O_p\Bigg\{\sqrt{nh^{d_A}}\left(\int_{\mathcal{X}\times\mathcal{M}} (\hat{\lambda}(a', M_i, X_i) - \lambda(a', M_i, X_i))^2 f(M_i, X_i)dM_i dX_i\right)^{1/2}$$
$$\left(\int_{\mathcal{X}\times\mathcal{M}} (\hat{\gamma}(a, M_i, X_i) - \gamma(a, M_i, X_i))^2 f(M_i, X_i)dM_i dX_i\right)^{1/2}\Bigg\}$$

$$\xrightarrow{p} 0$$

holds uniformly over $a', a \in \mathcal{A}_0$ by Assumption 6. The last inequality is obtained from Cauchy-Schwartz inequality and Assumption 5. Hence,

$$\sup_{a',a\in\mathcal{A}_0} \sqrt{nh^{d_A}}\mathbb{E}[\hat{g}(a',a)] = o_p(1) \Rightarrow \sup_{a',a\in\mathcal{A}_0} \mathbb{E}[\hat{g}(a',a)] = o_p\left(1/\sqrt{nh^{d_A}}\right)$$

□

Lemma 4 *For any $\epsilon > 0$, there exists a positive constant C such that $P(\mathcal{B}_n(C)) \geq 1 - \epsilon$ for n large enough, where $\mathcal{B}_n(C) = \cap_{l=1}^{L}\mathcal{B}_{ln}(C)$ and*

$$\mathcal{B}_{ln}(C) = \{\hat{\gamma}_\ell, \hat{\lambda}_\ell \colon \sup_{a\in\mathcal{A}_0, i\in I_\ell} |\hat{\gamma}_l(a, M_i, X_i) - \gamma(a, M_i, X_i)| \leq CA_{1n},$$
$$\sup_{a\in\mathcal{A}_0, i\in I_\ell} |\hat{\lambda}_l(a, M_i, X_i) - \lambda(a, M_i, X_i)| \leq CA_{2n}\}.$$

Proof: By Assumption 6 and the definition of O_p, for any $\epsilon > 0$, there exists $C > 0$ such that

$$P(\sup_{a\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}} |\hat{\gamma}_l(a, M_i, X_i) - \gamma(a, M_i, X_i)| > CA_{1n}) < \epsilon/(2L) \text{ and}$$

$$P(\sup_{a\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}} |\hat{\lambda}_l(a, M_i, X_i) - \lambda(a, M_i, X_i)| > CA_{2n}) < \epsilon/(2L)$$

for n large enough. Because

$$\begin{aligned}1 - P(\mathcal{B}_{ln}(C)) =&P(\sup_{a\in\mathcal{A}_0, i\in I_\ell}|\hat{\gamma}_l(a,M_i,X_i) - \gamma(a,M_i,X_i)| > CA_{1n},\\ &\sup_{a\in\mathcal{A}_0, i\in I_\ell}|\hat{\lambda}_l(a,M_i,X_i) - \lambda(a,M_i,X_i)| > CA_{2n})\\ \leq&P(\sup_{a\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}}|\hat{\gamma}_l(a,M_i,X_i) - \gamma(a,M_i,X_i)| > CA_{1n}),\\ &\sup_{a\in\mathcal{A}_0,(M_i,X_i)\in\mathcal{M}\times\mathcal{X}}|\hat{\lambda}_l(a,M_i,X_i) - \lambda(a,M_i,X_i)| > CA_{2n})\\ <&\epsilon/L,\end{aligned}$$

there is $P(\mathcal{B}_{ln}(C)) \geq 1 - \epsilon/L$. As a result,

$$\begin{aligned}&P(\mathcal{B}_n(C)) = P(\cap_{l=1}^L \mathcal{B}_{ln}(C))\\ \geq&\sum_{l=1}^L P(\mathcal{B}_{ln}(C)) - L + 1 \geq L - L\times\epsilon/L - L + 1 = 1 - \epsilon\end{aligned}$$

□

Lemma 5 *Given $\eta_n > 0$ and $\mathcal{B}_{ln}$ defined in lemma 4, the following equality holds*

$$P(n^{-1}W_{i\ell}(a',a) > \eta_n, \mathcal{B}_{ln}(C)) = \mathbb{E}[P(n^{-1}W_{i\ell}(a',a) > \eta_n \mid O_\ell^c)\mathbb{I}(\mathcal{B}_{ln}(C))]$$

Proof:

$$\begin{aligned}P(n^{-1}W_{i\ell}(a',a) > \eta_n, \mathcal{B}_{ln}(C)) &= P(n^{-1}W_{i\ell}(a',a) > \eta_n \mid \mathcal{B}_{ln}(C))P(\mathcal{B}_{ln}(C))\\ &= P(n^{-1}W_{i\ell}(a',a) > \eta_n \mid O_\ell^c)P(\mathcal{B}_{ln}(C))\\ &= \mathbb{E}[P(n^{-1}W_{i\ell}(a',a) > \eta_n \mid O_\ell^c)\mathbb{I}(\mathcal{B}_{ln}(C))]\end{aligned}$$

□

Lemma 6 *Given $\eta_n > 0$, for any positive sequence a_n and a random variable W satisfying $a_n|W| \leq 1/2$ and $\mathbb{E}(W) = 0$, there is*

$$P(W > \eta_n) \leq \exp(\mathbb{E}[a_n^2 W^2])/\exp(a_n\eta_n).$$

Proof:

$$\begin{aligned}&P(W > \eta_n) = P(a_nW > a_n\eta_n)\\ \leq&\mathbb{E}[\exp(a_nW)]/\exp(a_n\eta_n)\\ \leq&\exp(-a_n\eta_n)\left(1 + \mathbb{E}[a_n^2W^2]\right)\\ \leq&\exp(-a_n\eta_n)\exp(\mathbb{E}[a_n^2W^2])\end{aligned}$$

The first inequality comes from Markov inequality under $\exp(\cdot)$ being a monotonically increasing non-negative function. The second inequality is by $\exp(x) \leq 1+x+x^2$ for $|x| \leq 1/2$ taking $x = a_n W$ and $\mathbb{E}[a_n W] = 0$. The third inequality is from $1 + x \leq \exp(x)$ for $x \geq 0$.

□

Lemma 7 *For n large enough and $A_n = A_{1n}A_{2n}$ where A_{1n} and A_{2n} are defined in Assumption 6, there exists a positive constant c_1 such that $\mathbb{E}[W_{i\ell}(a', a)^2 \mid Z_c^l] \leq c_1 h^{-d_A} A_n^2$ for $a', a \in \mathcal{A}_0$.*

Proof: Write $O_i = (M_i, A_i, X_i)$, by $f(O_i|O_\ell^c) = f(O_i)$ for $i \in I_\ell$, there are the first and second order conditional moments bounded as below,

$$
\begin{aligned}
&\mathbb{E}[K_h(A_i - a')\Delta_{i\ell}(a', a)|O_\ell^c] = \mathbb{I}\{B_{ln}(C)\} \int \left\{ \prod_{j=1}^{d_A} \frac{k((A_{ij} - a'_j)/h)}{h} \right\} \Delta_{i\ell}(a', a) f(O_i) dO_i \\
&\leq \sup_{a', a \in \mathcal{A}_0} |\Delta_{i\ell}(a', a)| \sup_{O_i \in \mathcal{M} \times \mathcal{A}_0 \times \mathcal{X}} f(O_i) \int \prod_{j=1}^{d_A} k(u_j) du_j \\
&\leq \sup_{a', a \in \mathcal{A}_0} |\Delta_{i\ell}(a', a)| \sup_{O_i \in \mathcal{M} \times \mathcal{A}_0 \times \mathcal{X}} f(O_i) \prod_{j=1}^{d_A} \int k(u_j) du_j \\
&\leq C_1 A_n,
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}[(K_h(A_i - a')\Delta_{i\ell}(a', a))^2|O_\ell^c] = 1\{B_{ln}(C)\} \int \left\{ \prod_{j=1}^{d_A} \frac{k((A_{ij} - a'_j)/h)}{h} \right\}^2 \Delta_{i\ell}^2(a', a) f(O_i) dO_i \\
&\leq \left(\sup_{a', a \in \mathcal{A}_0} |\Delta_{i\ell}(a', a)| \right)^2 \left(\sup_{O_i \in \mathcal{M} \times \mathcal{A}_0 \times \mathcal{X}} f(O_i) \right) h^{-d_A} \prod_{j=1}^{d_A} \int k^2(u_j) du_j \\
&\leq C_2 h^{-d_A} A_n^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&E[W_{i\ell}(a', a)^2|O_\ell^c] = \mathbb{E}[\{K_h\Delta_{i\ell} - \mathbb{E}(K_h\Delta_{i\ell})\}^2|O_\ell^c] \\
&= \mathbb{E}[(K_h\Delta_{i\ell})^2|O_\ell^c] - 2\mathbb{E}[K_h\Delta_{i\ell}|O_\ell^c]\mathbb{E}[K_h\Delta_{i\ell}] + \mathbb{E}[K_h\Delta_{i\ell}]^2 \\
&\leq C_2 h^{-d_A} A_n^2 + o_p(1/\sqrt{nh^{d_A}})
\end{aligned}
$$

Since $o_p(1/\sqrt{nh^{d_A}})$ goes to zero, there always exists a constant c_1 large enough such that $\mathbb{E}[W_{i\ell}(a', a)^2 \mid Z_c^l] \leq c_1 h^{-d_A} A_n^2$ for large n. □

Because $\mathcal{A}_0$ is compact, $\mathcal{A}_0 \times \mathcal{A}_0$ is also compact. Thus, we can cover the range of (a', a) by a finite number M_n of cubes $\mathcal{C}_{k,n}$, each centered at $(a'_{k,n}, a_{k,n})$, $k = 1, \ldots, M_n$, and has length m_n. Since each cube has a volume of $m_n^{d_A}$ and $\mathcal{A}_0 \times \mathcal{A}_0$ has a finite volume, there is $M_n \propto m_n^{-d_A}$. To prove the uniform convergence of the target term, each of the following decomposed terms is bounded individually:

$$\begin{aligned}
&\sup_{a',a\in\mathcal{A}_0} |\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| \\
\leq &\max_{1\leq k\leq M_n} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} |\hat{g}(a',a) - \hat{g}(a'_{k,n}, a_{k,n})| && (9)\\
&+ \max_{1\leq k\leq M_n} |\hat{g}(a'_{k,n}, a_{k,n}) - \mathbb{E}[\hat{g}(a'_{k,n}, a_{k,n})]| && (10)\\
&+ \max_{1\leq k\leq M_n} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} |\mathbb{E}[\hat{g}(a'_{k,n}, a_{k,n})] - \mathbb{E}[\hat{g}(a',a)]|. && (11)
\end{aligned}$$

To begin with, we look at (10). By lemma 4, for any $\epsilon > 0$, there exists a $C > 0$ such that $P(\mathcal{B}_n(C)) \geq 1 - \epsilon/M_n$. Hence, $P(\mathcal{B}_n^c(C)) \leq \epsilon/M_n$ and

$$\begin{aligned}
&P(\max_{1\leq k\leq M_n} |\hat{g}(a'_{k,n}, a_{k,n}) - \mathbb{E}[\hat{g}(a'_{k,n}, a_{k,n})]| > \eta_n) \\
\leq &M_n \sup_{a,a'\in\mathcal{A}_0} P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n) \\
= &M_n \sup_{a,a'\in\mathcal{A}_0} P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n(C)) + M_n \sup_{a,a'\in\mathcal{A}_0} P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n^c(C)) \\
\leq &M_n \sup_{a,a'\in\mathcal{A}_0} P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n(C)) + M_n P(\mathcal{B}_n^c(C)) \\
\leq &M_n \sup_{a,a'\in\mathcal{A}_0} P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n(C)) + \epsilon. && (12)
\end{aligned}$$

So we simply have to bound $M_n \sup_{a,a'\in\mathcal{A}_0} P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n(C))$. Specifically, we examine $P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n(C))$ for $a, a' \in \mathcal{A}_0$.

$$\begin{aligned}
&P(|\hat{g}(a',a) - \mathbb{E}[\hat{g}(a',a)]| > \eta_n, \mathcal{B}_n(C)) \\
= &P(|n^{-1} \sum_{l=1}^{L} \sum_{i\in I_\ell} W_{i\ell}(a',a)| > \eta_n, \mathcal{B}_n(C)) \\
= &P(n^{-1} \sum_{l=1}^{L} \sum_{i\in I_\ell} W_{i\ell}(a',a) > \eta_n, \mathcal{B}_n(C)) + P(n^{-1} \sum_{l=1}^{L} \sum_{i\in I_\ell} W_{i\ell}(a',a) < -\eta_n, \mathcal{B}_n(C)) \\
\leq &\sum_{l=1}^{L} \sum_{i\in I_\ell} P(n^{-1} W_{i\ell}(a',a) > \eta_n, \mathcal{B}_{ln}(C)) + P(-n^{-1} W_{i\ell}(a',a) > \eta_n, \mathcal{B}_{ln}(C)) \\
= &\sum_{l=1}^{L} \sum_{i\in I_\ell} \mathbb{E}[P(n^{-1} W_{i\ell}(a',a) > \eta_n \mid O_\ell^c)\mathbb{I}(\mathcal{B}_{ln}(C))] + \mathbb{E}[P(-n^{-1} W_{i\ell}(a',a) > \eta_n \mid O_\ell^c)\mathbb{I}(\mathcal{B}_{ln}(C))] \\
\leq &2n \exp(-a_n \eta_n) \mathbb{E}[\exp(a_n^2 n^{-2} \mathbb{E}[W_{i\ell}^2(a',a) | O_\ell^c])] && (13)
\end{aligned}$$

The last equality comes from lemma 5. The last inequality is obtained from applying lemma 6 and its conditions are examined as follows. First, we have $\mathbb{E}[W_{i\ell}(a',a)] = 0$ from the definition of $W_{i\ell}(a',a)$. Second, let $A_n = A_{1n}A_{2n}$, when $\mathbb{I}(\mathcal{B}_{ln}(C)) = 1$, there is

$$\begin{aligned}&\sup_{a',a\in\mathcal{A}_0, i\in I_\ell} |W_{i\ell}(a',a)| \leq \sup_{a',a\in\mathcal{A}_0} |\Delta_{i\ell}(a',a)| \sup_{a'\in\mathcal{A}_0} K_h(A_i - a') + \sup_{a',a\in\mathcal{A}_0} \mathbb{E}[\hat{g}(a',a)]\\ \leq & C^2 A_n \sup_{a'\in\mathcal{A}_0} K_h(A_i - a') + o_p(1/\sqrt{nh^{d_A}}),\end{aligned}$$

so $W_{i\ell}(a',a)$ is bounded. Lastly, we choose $a_n = ln(n)\sqrt{nh^{d_t}}/A_n$, so $a_n/n \to 0$, and together with the boundedness of $W_{i\ell}$ on $\mathcal{B}_{ln}(C)$, there is $|a_n n^{-1} W_{i\ell}(t)| \leq 1/2$ for n large enough.

Let $\eta_n = c_2 A_n/\sqrt{nh^{d_t}}$, then $a_n\eta_n = ln(n)c_2 \to \infty$, and based on lemma 7, we have

$$\frac{a_n^2}{n^2}\mathbb{E}[W_{i\ell}(a',a)^2|O_\ell^c] \leq \frac{[ln(n)]^2 nh^{d_A}}{A_n^2}\frac{1}{n^2}c_1 h^{-d_A}A_n^2 = c_1\frac{[ln(n)]^2}{n},$$

which converges to zero by L'Hospital's rule. As a result, equation (12) is now bounded by

$$\begin{aligned}&P(\max_{1\leq k\leq M_n} |\hat{g}(a'_{k,n}, a_{k,n}) - \mathbb{E}[\hat{g}(a'_{k,n}, a_{k,n})]| > \eta_n)\\ \leq & M_n 2n \exp(-a_n\eta_n)\mathbb{E}[\exp(a_n^2 n^{-2}\mathbb{E}[W_{i\ell}(a',a)^2 \mid O_\ell^c])] + \epsilon\\ \leq & M_n 2n \exp\Bigg(-c_2 ln(n) + c_1\frac{[ln(n)]^2}{n}\Bigg) + \epsilon\\ \leq & 2M_n n^{c_1\frac{ln(n)}{n} - c_2 + 1} + \epsilon \leq 2\epsilon\end{aligned}$$

for n large enough by choosing c_2 such that $c_2 \geq 2$ and appropriate M_n. This indicates

$$\max_{1\leq k\leq M_n} |\hat{g}(t_{k,n}) - \mathbb{E}[\hat{g}(t_{k,n})]| = O_p(\eta_n)$$

, which is a rate of $o_p(1/\sqrt{nh^{d_t}})$ because $A_n \to 0$ by Assumption 6.

Next, we proof uniform convergence for (9). By the assumption that kernels $k(\cdot)$ are bounded, we have $K_h(A_i - a') = \frac{1}{h^{d_A}}\prod_{j=1}^{d_A} k(\frac{A_{ij}-a'_j}{h}) \leq c_3/h^{d_A}$. Then,

$$\begin{aligned}&\max_{1\leq k\leq M_n} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} |K_h(A_i - a')\Delta_{i\ell}(a',a) - K_h(A_i - a'_{k,n})\Delta_{i\ell}(a'_{k,n}, a_{k,n})|\\ \leq & c'_3 h^{-d_A} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} ||a - a_{k,n}|| \times ||a' - a'_{k,n}|| \leq c'_3 h^{-d_t} m_n^2\end{aligned}$$

where c'_3 is c_3 multiplied by the Lipschitz constant of $\Delta_{i\ell}$. By choosing $m_n = o_p((h^{d_t}/n)^{1/4})$, equation (9) is of rate $o_p(1/\sqrt{nh^{d_t}})$.

Previously the cubes were defined such that $M_n \propto 1/m_n^{d_t}$. So the choice of m_n leads to $M_n >> (nh^{-d_A})^{d_A/4}$. This is consistent with the proof of (10) above as long as c'_2 is large enough such that $M_n/n^{c_2-1} = o_p(1)$.

Lastly, for (11),

$$\max_{1\le k\le M_n} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} |\mathbb{E}[\hat{g}(a'_{k,n}, a_{k,n})] - \mathbb{E}[\hat{g}(a', a)]|$$
$$\le \max_{1\le k\le M_n} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} \mathbb{E}[|\hat{g}(a'_{k,n}, a_{k,n}) - \hat{g}(a', a)|]$$
$$\le \max_{1\le k\le M_n} \sup_{(a',a)\in(\mathcal{A}_0\times\mathcal{A}_0)\cap C_{k,n}} |\hat{g}(a'_{k,n}, a_{k,n}) - \hat{g}(a', a)| = o_p(1/\sqrt{nh^{d_t}})$$

The last equality comes from the rate of (9).

From (9), (10), and (11), there is

$$\sup_{a',a\in\mathcal{A}_0} |\hat{g}(a', a) - \mathbb{E}[\hat{g}(a', a)]| = o_p(1/\sqrt{nh^{d_t}}).$$

Our target term (8) can be written as

$$\sqrt{h^{d_A}/n}\sum_{l=1}^{L}\sum_{i\in I_\ell} K_h(A_i - a')\Delta_{i\ell}(a', a) = \sqrt{nh^{d_A}}\hat{g}(a', a)$$

and

$$\sup_{a',a\in\mathcal{A}_0} |\hat{g}(a', a)| \le \sup_{a',a\in\mathcal{A}_0} |\hat{g}(a', a) - \mathbb{E}[\hat{g}(a', a)]| + \sup_{a',a\in\mathcal{A}_0} |\mathbb{E}[\hat{g}(a', a)]| = o_p(1/\sqrt{nh^{d_A}}),$$

we know the target term is uniformly $o_p(1)$ over $a', a \in \mathcal{A}_0$.

Proof of (R1-1) and (R1-2)

We need to bound λ and $Y - \gamma$.